DEX-0291

CC

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068633 A2

(51) International Patent Classification[7]: C12N 15/09, C12Q 1/68, G01N 33/574, C07K 14/47

(21) International Application Number: PCT/US01/43612

(22) International Filing Date:
21 November 2001 (21.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/252,500 22 November 2000 (22.11.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/252,500 (CIP)
Filed on 22 November 2000 (22.11.2000)

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **CHEN, Sei-Yu** [—/US]; 160 Mira Street, Foster City, CA 94404 (US). **SUN, Yongming** [CN/US]; 551 Shoal Drive, Redwood City, CA 94065 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way, #14, San Jose, CA 95117 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/068633 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO LUNG SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic lung cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating lung cancer and non-cancerous disease states in lung, identifying lung tissue, monitoring and identifying and/or designing and antagonists of polypeptide of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered lung tissue for treatment and research.

39C

COMPOSITIONS AND METHODS RELATING TO LUNG SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application Serial No. 60/252,500 filed November 22, 2000, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic lung cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating lung cancer and non-cancerous disease states in lung, identifying lung tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered lung tissue for treatment and research.

BACKGROUND OF THE INVENTION

Throughout the last hundred years, the incidence of lung cancer has steadily increased, so much so that now in many countries, it is the most common cancer. In fact, lung cancer is the second most prevalent type of cancer for both men and women in the United States and is the most common cause of cancer death in both sexes. Lung cancer deaths have increased ten-fold in both men and women since 1930, primarily due to an increase in cigarette smoking, but also due to an increased exposure to arsenic, asbestos, chromates, chloromethyl ethers, nickel, polycyclic aromatic hydrocarbons and other agents. *See* Scott, Lung Cancer: A Guide to Diagnosis and Treatment, Addicus Books (2000) and Alberg *et al.*, in Kane *et al.* (eds.) Biology of Lung Cancer, pp. 11-52, Marcel Dekker, Inc. (1998). Lung cancer may result from a primary tumor originating in the

lung or a secondary tumor which has spread from another organ such as the bowel or breast. Although there are over a dozen types of lung cancer, over 90% fall into two categories: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). *See* Scott, *supra*. About 20-25% of all lung cancers are characterized as SCLC, while 70-80% are diagnosed as NSCLC. *Id.* A rare type of lung cancer is mesothelioma, which is generally caused by exposure to asbestos, and which affects the pleura of the lung. Lung cancer is usually diagnosed or screened for by chest x-ray, CAT scans, PET scans, or by sputum cytology. A diagnosis of lung cancer is usually confirmed by biopsy of the tissue. *Id.*

SCLC tumors are highly metastatic and grow quickly. By the time a patient has been diagnosed with SCLC, the cancer has usually already spread to other parts of the body, including lymph nodes, adrenals, liver, bone, brain and bone marrow. *See* Scott, *supra*; Van Houtte *et al.* (eds.), Progress and Perspective in the Treatment of Lung Cancer, Springer-Verlag (1999). Because the disease has usually spread to such an extent that surgery is not an option, the current treatment of choice is chemotherapy plus chest irradiation. *See* Van Houtte, *supra*. The stage of disease is a principal predictor of long-term survival. Less than 5% of patients with extensive disease that has spread beyond one lung and surrounding lymph nodes, live longer than two years. *Id.* However, the probability of five-year survival is three to four times higher if the disease is diagnosed and treated when it is still in a limited stage, i.e., not having spread beyond one lung. *Id.*

NSCLC is generally divided into three types: squamous cell carcinoma, adenocarcinoma and large cell carcinoma. Both squamous cell cancer and adenocarcinoma develop from the cells that line the airways; however, adenocarcinoma develops from the goblet cells that produce mucus. Large cell lung cancer has been thus named because the cells look large and rounded when viewed microscopically, and generally are considered relatively undifferentiated. *See* Yesner, Atlas of Lung Cancer, Lippincott-Raven (1998).

Secondary lung cancer is a cancer initiated elsewhere in the body that has spread to the lungs. Cancers that metastasize to the lung include, but are not limited to, breast cancer, melanoma, colon cancer and Hodgkin's lymphoma. Treatment for secondary lung cancer may depend upon the source of the original cancer. In other words, a lung

cancer that originated from breast cancer may be more responsive to breast cancer treatments and a lung cancer that originated from the colon cancer may be more responsive to colon cancer treatments.

The stage of a cancer indicates how far it has spread and is an important indicator of the prognosis. In addition, staging is important because treatment is often decided according to the stage of a cancer. SCLC is divided into two stages: limited disease, *i.e.*, cancer that can only be seen in one lung and in nearby lymph nodes; and extensive disease, *i.e.*, cancer that has spread outside the lung to the chest or to other parts of the body. For most patients with SCLC, the disease has already progressed to lymph nodes or elsewhere in the body at the time of diagnosis. *See* Scott, *supra*. Even if spreading is not apparent on the scans, it is likely that some cancer cells may have spread away and traveled through the bloodstream or lymph system. In general, chemotherapy with or without radiotherapy is often the preferred treatment. The initial scans and tests done at first will be used later to see how well a patient is responding to treatment.

In contrast, non-small cell cancer may be divided into four stages. Stage I is highly localized cancer with no cancer in the lymph nodes. Stage II cancer has spread to the lymph nodes at the top of the affected lung. Stage III cancer has spread near to where the cancer started. This can be to the chest wall, the covering of the lung (pleura), the middle of the chest (mediastinum) or other lymph nodes. Stage IV cancer has spread to another part of the body. Stage I-III cancer is usually treated with surgery, with or without chemotherapy. Stage IV cancer is usually treated with chemotherapy and/or palliative care.

A number of chromosomal and genetic abnormalities have been observed in lung cancer. In NSCLC, chromosomal aberrations have been described on 3p, 9p, 11p, 15p and 17p, and chromosomal deletions have been seen on chromosomes 7, 11, 13 and 19. *See* Skarin (ed.), Multimodality Treatment of Lung Cancer, Marcel Dekker, Inc. (2000); Gemmill *et al.*, pp. 465-502, in Kane, *supra*; Bailey-Wilson *et al.*, pp. 53-98, in Kane, *supra*. Chromosomal abnormalities have been described on 1p, 3p, 5q, 6q, 8q, 13q and 17p in SCLC. *Id.* The loss of the short arm of chromosome 3p has also been seen in greater than 90% of SCLC tumors and approximately 50% of NSCLC tumors. *Id.*

A number of oncogenes and tumor suppressor genes have been implicated in lung cancer. *See* Mabry, pp. 391-412, in Kane, *supra* and Sclafani *et al.*, pp. 295-316, in

Kane, *supra*. In both SCLC and NSCLC, the p53 tumor suppressor gene is mutated in over 50% of lung cancers. *See* Yesner, *supra*. Another tumor suppressor gene, FHIT, which is found on chromosome 3p, is mutated by tobacco smoke. *Id.*; Skarin, *supra*. In addition, more than 95% of SCLCs and approximately 20-60% of NSCLCs have an absent or abnormal retinoblastoma (Rb) protein, another tumor suppressor gene. The *ras* oncogene (particularly K-*ras*) is mutated in 20-30% of NSCLC specimens and the c-*erbB2* oncogene is expressed in 18% of stage 2 NSCLC and 60% of stage 4 NSCLC specimens. *See* Van Houtte, *supra*. Other tumor suppressor genes that are found in a region of chromosome 9, specifically in the region of 9p21, are deleted in many cancer cells, including $p16^{INK4A}$ and $p15^{INK4B}$. *See* Bailey-Wilson, *supra*; Sclafani *et al.*, *supra*. These tumor suppressor genes may also be implicated in lung cancer pathogenesis.

In addition, many lung cancer cells produce growth factors that may act in an autocrine fashion on lung cancer cells. *See* Siegfried *et al.*, pp. 317-336, in Kane, *supra*; Moody, pp. 337-370, in Kane, *supra* and Heasley *et al.*, 371-390, in Kane, *supra*. In SCLC, many tumor cells produce gastrin-releasing peptide (GRP), which is a proliferative growth factor for these cells. *See* Skarin, *supra*. Many NSCLC tumors express epidermal growth factor (EGF) receptors, allowing NSCLC cells to proliferate in response to EGF. Insulin-like growth factor (IGF-I) is elevated in greater than 95% of SCLC and greater than 80% of NSCLC tumors; it is thought to function as an autocrine growth factor. *Id.* Finally, stem cell factor (SCF, also known as steel factor or kit ligand) and c-Kit (a proto-oncoprotein tyrosine kinase receptor for SCF) are both expressed at high levels in SCLC, and thus may form an autocrine loop that increases proliferation. *Id.*

Although the majority of lung cancer cases are attributable to cigarette smoking, most smokers do not develop lung cancer. Epidemiological evidence has suggested that susceptibility to lung cancer may be inherited in a Mendelian fashion, and thus have an inherited genetic component. Bailey-Wilson, *supra*. Thus, it is thought that certain allelic variants at some genetic loci may affect susceptibility to lung cancer. *Id.* One way to identify which allelic variants are likely to be involved in lung cancer susceptibility, as well as susceptibility to other diseases, is to look at allelic variants of genes that are highly expressed in lung.

The lung is susceptible to a number of other debilitating diseases as well, including, without limitation, emphysema, pneumonia, cystic fibrosis and asthma. *See* Stockley (ed.), Molecular Biology of the Lung, Volume I: Emphysema and Infection, Birkhauser Verlag (1999), hereafter Stockley I, and Stockley (ed.), Molecular Biology of the Lung, Volume II: Asthma and Cancer, Birkhauser Verlag (1999), hereafter Stockley II. The cause of many these disorders is still not well understood and there are few, if any, good treatment options for many of these noncancerous lung disorders. Thus, there also remains a need for understanding of various noncancerous lung disorders and for identify treatments for these diseases.

The development and differentiation of the lung tissue during embryonic development is also very important. All of the epithelial cells of the respiratory tract, including those of the lung and bronchi, are derived from the primitive endodermal cells that line the embryonic outpouching. *See* Yesner, *supra*. During embryonic development, multipotent endodermal stem cells differentiate into many different types of specialized cells, which include ciliated cells for moving inhaled particles, goblet cells for producing mucus, Kulchitsky's cells for endocrine function, and Clara cells and type II pneumocytes for secreting surfactant protein. *Id.* Improper development and differentiation may cause respiratory disorders and distress in infants, particularly in premature infants, whose lungs cannot produce sufficient surfactant when they are born. Further, some lung cancer cells, particularly small cell carcinomas, appear multipotent, and can spontaneously differentiate into a number of cell types, including small cell carcinoma, adenocarcinoma and squamous cell carcinoma. *Id.* Thus, a better understanding of lung development and differentiation may help facilitate understanding of lung cancer initiation and progression.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop lung cancer, for diagnosing lung cancer, for monitoring the progression of the disease, for staging the lung cancer, for determining whether the lung cancer has metastasized and for imaging the lung cancer. There is also a need for better treatment of lung cancer. There is also a great need for diagnosing and treating noncancerous lung disorders such as emphysema, pneumonia, lung infection, pulmonary fibrosis, cystic fibrosis and asthma. There is also a need for compositions and methods of using compositions that are capable of identifying lung tissue for forensic

purposes and for determining whether a particular cell or tissue exhibits lung-specific characteristics.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat lung cancer and non-cancerous disease states in lung; identify and monitor lung tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered lung tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to lung cells, lung tissue and/or the lung organ. These lung specific nucleic acids (LSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the LSNA is genomic DNA, then the LSNA is a lung specific gene (LSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to lung. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 165 through 284. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 164. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding an LSP, or that selectively hybridize or exhibit substantial sequence similarity to an LSNA, as well as allelic variants of a nucleic acid molecule encoding an LSP, and allelic variants of an LSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes an LSP or that comprises a part of a nucleic acid sequence of an LSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of an LSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the

transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of an LSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of an LSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of an LSNA.

Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is an LSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of an LSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating lung cancer and non-cancerous disease states in lung. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring lung tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered lung tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat lung cancer and non-cancerous disease states in lung. The invention provides methods of using the polypeptides of the invention to identify and/or monitor lung tissue, and to produce engineered lung tissue.

The agonists and antagonists of the instant invention may be used to treat lung cancer and non-cancerous disease states in lung and to produce engineered lung tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and

pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround

the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid

molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs

typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology"

interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence

hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41$ (fraction G + C) – 0.63 (% formamide) - (600/l)

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify

nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula: $T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$, wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAS. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one

exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding an LSP or is an LSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455

(1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable

of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises an LSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation,

covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$--, and –CH_2SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al., Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino

acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed.,

Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as -, -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, -N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino

acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.,* Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an $F(ab')_2$ fragment is a bivalent fragment comprising two Fab

fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.,* Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.,* Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g.*, Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany

it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "lung specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the lung as compared to other tissues in the body. In a preferred embodiment, a "lung specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "lung specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the lung or to lung cells or tissue or that are derived from such nucleic acid molecules. These isolated lung specific nucleic acids (LSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to lung, a lung-specific polypeptide (LSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 165 through 284. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through164.

An LSNA may be derived from a human or from another animal. In a preferred embodiment, the LSNA is derived from a human or other mammal. In a more preferred embodiment, the LSNA is derived from a human or other primate. In an even more preferred embodiment, the LSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding an LSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode an LSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes an LSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 165 through 284. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 164.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an LSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an LSP under moderate stringency conditions. In a more preferred

embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an LSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 165 through 284. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 164. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding an LSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human LSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 165 through 284. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding an LSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 165 through 284, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding an LSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding an LSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to an LSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 164. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity

with an LSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 164, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with an LSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with an LSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to an LSNA or to a nucleic acid molecule encoding an LSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the LSNA or the nucleic acid molecule encoding an LSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 165 through 284 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 164. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the LSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is

experimentally produced by directed mutation of an LSNA. Further, the substantially similar nucleic acid molecule may or may not be an LSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is an LSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of an LSNA or a nucleic acid encoding an LSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes an LSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is an LSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 164. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is an LSP. However, in a preferred embodiment, the part encodes an LSP. In one aspect, the invention comprises a part of an LSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to an LSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of an LSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes an LSP. A part comprises at least 10 nucleotides, more preferably at least 15,

17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as -^{32}P-dATP, -^{32}P-dCTP, -^{32}P-dGTP, -^{32}P-dTTP, -^{32}P-3'dATP, -^{32}P-ATP, -^{32}P-CTP, -^{32}P-GTP, -^{32}P-UTP, -^{35}S-dATP, α-^{35}S-GTP, α-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3'

hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.*, Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.),

Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. *See* Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S.

Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases

do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably,

detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of an LSNA, such as deletions, insertions, translocations, and duplications of the LSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify LSNA in, and isolate LSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^+- selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to LSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding an LSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 165 through 284. In another preferred embodiment, the probe or primer is derived from an LSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 164.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular

Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention

can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides

encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their

derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda* , *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-

transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of

the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast _-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early

gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the LSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PltetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to

be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from

those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,*. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.,* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.,* Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable

expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,

and acylation, and it is an aspect of the present invention to provide LSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database http://www.abrf.org/ABRF/Research Committees/deltamass/deltamass.html (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) and http://www.glycosuite.com/ (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999) and http://www.cbs.dtu.dk/databases/OGLYCBASE/ (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1):237-239 (1999) and http://www.cbs.dtu.dk/databases/PhosphoBase/ (accessed October 19, 2001); or http://pir.georgetown.edu/pirwww/search/textresid.html (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur

in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors

typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from lung are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human lung cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See, e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli,* in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^{+} or K^{+}, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols

(BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent,

FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al., Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a lung specific polypeptide (LSP). In an even more preferred embodiment, the polypeptide

is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 165 through 284. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of an LSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 165 through 284. A polypeptide that comprises only a fragment of an entire LSP may or may not be a polypeptide that is also an LSP. For instance, a full-length polypeptide may be lung-specific, while a fragment thereof may be found in other tissues as well as in lung. A polypeptide that is not an LSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-LSP antibodies. However, in a preferred embodiment, the part or fragment is an LSP. Methods of determining whether a polypeptide is an LSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, an LSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably an LSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably an LSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence

of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be lung-specific. In a preferred embodiment, the mutein is lung-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 165 through 284. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is lung-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*,

Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to an LSP. In an even more preferred embodiment, the polypeptide is homologous to an LSP selected from the group having an amino acid sequence of SEQ ID NO: 165 through 284. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to an LSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 165 through 284. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to an LSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to an LSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the LSNA is selected from the group consisting of SEQ ID NO: 1 through 164. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes an LSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred

embodiment, the LSP is selected from the group consisting of SEQ ID NO: 165 through 284.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 165 through 284. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the LSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of an LSP. Further, the homologous protein may or may not encode polypeptide that is an LSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is an LSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding an LSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 165 through 284. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 164.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an LSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 165 through 284, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514,

Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-LSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated

with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an LSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 165 through 284. In a preferred embodiment, the analog is one that comprises one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to an LSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$-- and –CH_2SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of an LSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g.,* Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyl chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyl quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-

aminobenzoyl)-β-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is an LSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 165 through 284, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic

acid sequence of SEQ ID NO: 1 through 164, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 164.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a

glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbrizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in Drosophila with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast

GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β-galactosidase, biotin trpE, protein A, β-lactamase, -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast _ mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See*, *e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the LSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize LSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly LSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for

detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of LSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of LSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.,* Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is an LSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 165 through 284, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a LSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of a LSP may be indicative of cancer. Differential degradation of the C or N-terminus of a LSP may also be a marker or target for anticancer therapy. For example, a LSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-LSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human lung.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about $5 \times 10^{-}$

^{7} M, 1 x 10^{-7} M, with affinities and avidities of at least 1 x 10^{-8} M, 5 x 10^{-9} M, 1 x 10^{-10} M and up to 1 X 10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are hereby incorporated in their entirety. In such cases, as with the transgenic human-

antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. *See* Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in*

Biotechnol. 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.,* Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.,* Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et*

al., *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); Limonta et al., *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, $F(ab)'_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or

one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or

transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic

compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g.*, Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I.

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb, ^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding an LSP. In a preferred embodiment, the LSP comprises an amino

acid sequence selected from SEQ ID NO: 165 through 284, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise an LSNA of the invention, preferably an LSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 164, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human LSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.*, Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.,* Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.,* Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.,* Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.*, Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.,* Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,* Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.*, a

nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.,* Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.*, Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC

compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.,* U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 164 and SEQ ID NO: 165 through 284 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set

representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Lung Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by

comparing expression of an LSNA or an LSP in a human patient that has or may have lung cancer, or who is at risk of developing lung cancer, with the expression of an LSNA or an LSP in a normal human control. For purposes of the present invention, "expression of an LSNA" or "LSNA expression" means the quantity of LSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of an LSP" or "LSP expression" means the amount of LSP that can be measured by any method known in the art or the level of translation of an LSG LSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing lung cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of LSNA or LSP in cells, tissues, organs or bodily fluids compared with levels of LSNA or LSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of an LSNA or LSP in the patient versus the normal human control is associated with the presence of lung cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing lung cancer in a patient by analyzing changes in the structure of the mRNA of an LSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing lung cancer in a patient by analyzing changes in an LSP compared to an LSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the LSP or subcellular LSP localization.

In a preferred embodiment, the expression of an LSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 165 through 284, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the LSNA expression that is measured is the level of expression of an LSNA mRNA selected from SEQ ID NO: 1 through 164, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. LSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by

Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. LSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of an LSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, LSNA expression may be compared to a known control, such as normal lung nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of an LSP is measured by determining the level of an LSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 165 through 284, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of LSNA or LSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of lung cancer. The expression level of an LSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the LSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the LSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to an LSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-LSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the LSP will bind to the anti-LSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-LSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the LSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an LSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure LSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-LSP antibody is attached to a solid support and an allocated amount of a labeled LSP and a sample of interest are incubated with the solid support. The amount of labeled LSP detected which is attached to the solid support can be correlated to the quantity of an LSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of an LSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other

mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more LSNAs of interest. In this approach, all or a portion of one or more LSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of LSNA or LSP includes, without limitation, lung tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, lung cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary lung cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in LSNAs or LSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of an LSNA or LSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other LSNA or LSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular LSNA or LSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more LSNAs and/or LSPs in a sample from a patient suspected of having lung cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of an LSNA and/or LSP and then ascertaining whether the patient has lung cancer from the expression level of the LSNA or LSP. In general, if high expression relative to a control of an LSNA or LSP is indicative of lung cancer, a diagnostic assay is considered positive if the level of expression of the LSNA or LSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an LSNA or LSP is indicative of lung cancer, a diagnostic assay is considered positive if the level of expression of the LSNA or LSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether lung cancer has metastasized in a patient. One may identify whether the lung cancer has metastasized by measuring the expression levels and/or structural alterations of one or more LSNAs and/or LSPs in a variety of tissues. The presence of an LSNA or LSP in a certain tissue

at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of an LSNA or LSP is associated with lung cancer. Similarly, the presence of an LSNA or LSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of an LSNA or LSP is associated with lung cancer. Further, the presence of a structurally altered LSNA or LSP that is associated with lung cancer is also indicative of metastasis.

In general, if high expression relative to a control of an LSNA or LSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the LSNA or LSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an LSNA or LSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the LSNA or LSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The LSNA or LSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with lung cancers or other lung related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of lung disorders.

Staging

The invention also provides a method of staging lung cancer in a human patient. The method comprises identifying a human patient having lung cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more LSNAs or LSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more LSNAs or LSPs is determined for each stage to obtain a standard expression level for each LSNA and LSP. Then, the LSNA or LSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The LSNA or LSP expression levels from the patient are then compared to the

standard expression level. By comparing the expression level of the LSNAs and LSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of an LSNA or LSP to determine the stage of a lung cancer.

Monitoring

Further provided is a method of monitoring lung cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the lung cancer. The method comprises identifying a human patient that one wants to monitor for lung cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more LSNAs or LSPs, and comparing the LSNA or LSP levels over time to those LSNA or LSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in an LSNA or LSP that are associated with lung cancer.

If increased expression of an LSNA or LSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of an LSNA or LSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of an LSNA or LSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of an LSNA or LSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of LSNAs or LSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of lung cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of an LSNA and/or LSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more LSNAs and/or LSPs are detected. The presence of higher (or lower) LSNA or LSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly lung cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more LSNAs and/or LSPs of the invention can also be monitored by analyzing levels of expression of the LSNAs and/or LSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in an LSG, thereby determining if a human with the genetic lesion is susceptible to developing lung cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing lung cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the LSGs of this invention, a chromosomal rearrangement of LSG, an aberrant modification of LSG (such as of the methylation pattern of the genomic DNA), or allelic loss of an LSG. Methods to detect such lesions in the LSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Lung Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more LSNAs and/or LSPs in a sample from a patient suspected of having or known to have a noncancerous lung disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of an LSNA and/or LSP, comparing the expression level or structural alteration of the LSNA or LSP to a normal lung control,

and then ascertaining whether the patient has a noncancerous lung disease. In general, if high expression relative to a control of an LSNA or LSP is indicative of a particular noncancerous lung disease, a diagnostic assay is considered positive if the level of expression of the LSNA or LSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an LSNA or LSP is indicative of a noncancerous lung disease, a diagnostic assay is considered positive if the level of expression of the LSNA or LSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether an LSNA and/or LSP is associated with a particular noncancerous lung disease by obtaining lung tissue from a patient having a noncancerous lung disease of interest and determining which LSNAs and/or LSPs are expressed in the tissue at either a higher or a lower level than in normal lung tissue. In another embodiment, one may determine whether an LSNA or LSP exhibits structural alterations in a particular noncancerous lung disease state by obtaining lung tissue from a patient having a noncancerous lung disease of interest and determining the structural alterations in one or more LSNAs and/or LSPs relative to normal lung tissue.

Methods for Identifying Lung Tissue

In another aspect, the invention provides methods for identifying lung tissue. These methods are particularly useful in, *e.g.*, forensic science, lung cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is lung tissue or has lung tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising lung tissue or having lung tissue-like characteristics, determining whether the sample expresses one or more LSNAs and/or LSPs, and, if the sample expresses one or more LSNAs and/or LSPs, concluding that the sample comprises lung tissue. In a preferred embodiment, the LSNA encodes a

polypeptide having an amino acid sequence selected from SEQ ID NO: 165 through 284, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the LSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 164, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses an LSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether an LSP is expressed. Determining whether a sample expresses an LSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the LSP has an amino acid sequence selected from SEQ ID NO: 165 through 284, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two LSNAs and/or LSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five LSNAs and/or LSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is lung tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into lung tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new lung tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Lung Tissue

In another aspect, the invention provides methods for producing engineered lung tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing an LSNA or an LSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of lung tissue cells. In a preferred

embodiment, the cells are pluripotent. As is well-known in the art, normal lung tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered lung tissue or cells comprises one of these cell types. In another embodiment, the engineered lung tissue or cells comprises more than one lung cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the lung cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more LSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode LSPs having amino acid sequences selected from SEQ ID NO: 165 through 284, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 164, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, an LSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial lung tissue may be used to treat patients who have lost some or all of their lung function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises an LSNA or part thereof. In a more preferred embodiment, the LSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 164, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises an LSP or fragment thereof. In a more preferred embodiment, the LSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 165 through 284, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the

pharmaceutical composition comprises an anti-LSP antibody, preferably an antibody that specifically binds to an LSP having an amino acid that is selected from the group consisting of SEQ ID NO: 165 through 284, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot

injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts

tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example LSP polypeptide, fusion protein, or fragments thereof, antibodies specific for LSP, agonists, antagonists or inhibitors of LSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age,

weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention , *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of lung function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication

incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV) , for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of an LSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of an LSP are administered, for example, to complement a deficiency in the native LSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes an LSP having the amino acid sequence of SEQ ID NO: 165 through 284, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express an LSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in LSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode an LSP having the amino acid sequence of SEQ ID NO: 165 through 284, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of an LSG antisense nucleic acid, are administered to downregulate transcription and/or translation of an LSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of an LSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to LSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the LSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.,* Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding an LSP, preferably an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 164, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an LSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant LSP defect.

Protein compositions are administered, for example, to complement a deficiency in native LSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to LSP. The immune response can

be used to modulate activity of LSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate LSP.

In a preferred embodiment, the polypeptide is an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 164, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of LSP, or to target therapeutic agents to sites of LSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to an LSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 164, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to an LSP or have a modulatory effect on the expression or activity of an LSP. Modulators which decrease the expression or activity of LSP (antagonists) are believed to be useful in treating lung cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of an LSP can also be designed, synthesized and tested for use in the imaging and treatment of lung cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the LSPs identified herein. Molecules identified in the library as being capable of binding to an LSP are key candidates for

further evaluation for use in the treatment of lung cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of an LSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of LSP is administered. Antagonists of LSP can be produced using methods generally known in the art. In particular, purified LSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of an LSP.

In other embodiments a pharmaceutical composition comprising an agonist of an LSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an LSP comprising an amino acid sequence of SEQ ID NO: 165 through 284, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an LSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 164, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Lung Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the lung or to specific cells in the lung. In a preferred embodiment, an anti-LSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if lung tissue needs to be selectively destroyed. This would be useful for targeting and killing lung cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting lung cell function.

In another embodiment, an anti-LSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring lung function, identifying lung cancer tumors, and identifying noncancerous lung diseases.

EXAMPLES

Example 1: Gene Expression analysis

LSGs were identified by mRNA subtraction analysis using standard methods. The sequences were extended using GeneBank sequences, Incyte's proprietary database. From the nucleotide sequences, predicted amino acid sequences were prepared. DEX0291_1, DEX0291_2 correspond to SEQ ID NO.1, 2 etc. DEX0134 was the parent sequence found in the mRNA subtractions.

DEX0291_1	DEX0134_1	DEX0291_165
DEX0291_2	flex DEX0134_1	
DEX0291_3	DEX0134_2	DEX0291_166
DEX0291_4	flex DEX0134_2	
DEX0291_5	DEX0134_3	DEX0291_167
DEX0291_6	flex DEX0134_3	DEX0291_168
DEX0291_7	DEX0134_4	DEX0291_169
DEX0291_8	flex DEX0134_4	DEX0291_170
DEX0291_9	DEX0134_5	DEX0291_171
DEX0291_10	flex DEX0134_5	
DEX0291_11	DEX0134_6	DEX0291_172
DEX0291_12	flex DEX0134_6	
DEX0291_13	DEX0134_7	DEX0291_173
DEX0291_14	flex DEX0134_7	
DEX0291_15	DEX0134_8	DEX0291_174
DEX0291_16	flex DEX0134_8	DEX0291_175
DEX0291_17	DEX0134_9	
DEX0291_18	DEX0134_10	DEX0291_176
DEX0291_19	DEX0134_11	DEX0291_177
DEX0291_20	flex DEX0134_11	
DEX0291_21	DEX0134_12	DEX0291_178
DEX0291_22	flex DEX0134_12	DEX0291_179

DEX0291_23 DEX0134_13 DEX0291_180

DEX0291_24 DEX0134_14 DEX0291_181

DEX0291_25 DEX0134_15 DEX0291_182

DEX0291_26 flex DEX0134_15 DEX0291_183

DEX0291_27 DEX0134_16 DEX0291_184

DEX0291_28 flex DEX0134_16 DEX0291_185

DEX0291_29 DEX0134_17 DEX0291_186

DEX0291_30 flex DEX0134_17 DEX0291_187

DEX0291_31 DEX0134_18 DEX0291_188

DEX0291_32 DEX0134_19 DEX0291_189

DEX0291_33 DEX0134_20 DEX0291_190

DEX0291_34 flex DEX0134_20 DEX0291_191

DEX0291_35 DEX0134_21 DEX0291_192

DEX0291_36 flex DEX0134_21

DEX0291_37 DEX0134_22 DEX0291_193

DEX0291_38 flex DEX0134_22 DEX0291_194

DEX0291_39 DEX0134_23

DEX0291_40 DEX0134_24 DEX0291_195

DEX0291_41 DEX0134_25

DEX0291_42 DEX0134_27 DEX0291_196

DEX0291_43 flex DEX0134_27

DEX0291_44 DEX0134_28

DEX0291_45 DEX0134_29 DEX0291_197

DEX0291_46 flex DEX0134_29 DEX0291_198

DEX0291_47 DEX0134_30 DEX0291_199

DEX0291_48 flex DEX0134_30 DEX0291_200

DEX0291_49 DEX0134_31 DEX0291_201

DEX0291_50 flex DEX0134_31

DEX0291_51 DEX0134_32 DEX0291_202

	DEX0291_52	DEX0134_33	
	DEX0291_53	DEX0134_34	DEX0291_203
	DEX0291_54	flex DEX0134_34	DEX0291_204
	DEX0291_55	DEX0134_35	DEX0291_205
5	DEX0291_56	flex DEX0134_35	DEX0291_206
	DEX0291_57	DEX0134_36	DEX0291_207
	DEX0291_58	flex DEX0134_36	DEX0291_208
	DEX0291_59	DEX0134_37	DEX0291_209
	DEX0291_60	flex DEX0134_37	DEX0291_210
10	DEX0291_61	DEX0134_38	DEX0291_211
	DEX0291_62	flex DEX0134_38	DEX0291_212
	DEX0291_63	DEX0134_39	DEX0291_213
	DEX0291_64	flex DEX0134_39	
	DEX0291_65	DEX0134_40	DEX0291_214
15	DEX0291_66	flex DEX0134_40	
	DEX0291_67	DEX0134_41	DEX0291_215
	DEX0291_68	flex DEX0134_41	
	DEX0291_69	DEX0134_42	DEX0291_216
	DEX0291_70	flex DEX0134_42	
20	DEX0291_71	DEX0134_43	
	DEX0291_72	DEX0134_44	DEX0291_217
	DEX0291_73	flex DEX0134_44	DEX0291_218
	DEX0291_74	DEX0134_46	
	DEX0291_75	flex DEX0134_46	
25	DEX0291_76	DEX0134_47	DEX0291_219
	DEX0291_77	flex DEX0134_47	DEX0291_220
	DEX0291_78	DEX0134_48	DEX0291_221
	DEX0291_79	flex DEX0134_48	DEX0291_222
	DEX0291_80	DEX0134_49	DEX0291_223

DEX0291_81 flex DEX0134_49 DEX0291_224

DEX0291_82 DEX0134_51 DEX0291_225

DEX0291_83 flex DEX0134_51

DEX0291_84 DEX0134_52 DEX0291_226

DEX0291_85 flex DEX0134_52

DEX0291_86 DEX0134_53 DEX0291_227

DEX0291_87 flex DEX0134_53 DEX0291_228

DEX0291_88 DEX0134_54 DEX0291_229

DEX0291_89 flex DEX0134_54 DEX0291_230

DEX0291_90 DEX0134_55 DEX0291_231

DEX0291_91 flex DEX0134_55

DEX0291_92 DEX0134_56 DEX0291_232

DEX0291_93 flex DEX0134_56 DEX0291_233

DEX0291_94 DEX0134_57 DEX0291_234

DEX0291_95 DEX0134_58 DEX0291_235

DEX0291_96 flex DEX0134_58

DEX0291_97 DEX0134_60 DEX0291_236

DEX0291_98 flex DEX0134_60

DEX0291_99 DEX0134_61 DEX0291_237

DEX0291_100 flex DEX0134_61

DEX0291_101 DEX0134_62 DEX0291_238

DEX0291_102 flex DEX0134_62 DEX0291_239

DEX0291_103 DEX0134_63 DEX0291_240

DEX0291_104 DEX0134_64 DEX0291_241

DEX0291_105 flex DEX0134_64

DEX0291_106 DEX0134_65 DEX0291_242

DEX0291_107 flex DEX0134_65 DEX0291_243

DEX0291_108 DEX0134_66 DEX0291_244

DEX0291_109 flex DEX0134_66

DEX0291_110	DEX0134_67		
DEX0291_111	flex DEX0134_67	DEX0291_245	
DEX0291_112	DEX0134_68	DEX0291_246	
DEX0291_113	flex DEX0134_68	DEX0291_247	
DEX0291_114	DEX0134_69	DEX0291_248	
DEX0291_115	flex DEX0134_69	DEX0291_249	
DEX0291_116	DEX0134_70	DEX0291_250	
DEX0291_117	flex DEX0134_70	DEX0291_251	
DEX0291_118	DEX0134_71	DEX0291_252	
DEX0291_119	DEX0134_72	DEX0291_253	
DEX0291_120	flex DEX0134_72		
DEX0291_121	DEX0134_73	DEX0291_254	
DEX0291_122	flex DEX0134_73	DEX0291_255	
DEX0291_123	DEX0134_74	DEX0291_256	
DEX0291_124	flex DEX0134_74		
DEX0291_125	DEX0134_75	DEX0291_257	
DEX0291_126	flex DEX0134_75		
DEX0291_127	DEX0134_76	DEX0291_258	
DEX0291_128	flex DEX0134_76	DEX0291_259	
DEX0291_129	DEX0134_77	DEX0291_260	
DEX0291_130	flex DEX0134_77		
DEX0291_131	DEX0134_78	DEX0291_261	
DEX0291_132	flex DEX0134_78		
DEX0291_133	DEX0134_79	DEX0291_262	
DEX0291_134	DEX0134_80	DEX0291_263	
DEX0291_135	flex DEX0134_80	DEX0291_264	
DEX0291_136	DEX0134_81	DEX0291_265	
DEX0291_137	DEX0134_82	DEX0291_266	
DEX0291_138	DEX0134_83	DEX0291_267	

DEX0291_139 flex DEX0134_83

DEX0291_140 DEX0134_84 DEX0291_268

DEX0291_141 flex DEX0134_84

DEX0291_142 DEX0134_85 DEX0291_269

DEX0291_143 DEX0134_86 DEX0291_270

DEX0291_144 flex DEX0134_86

DEX0291_145 DEX0134_87

DEX0291_146 DEX0134_88 DEX0291_271

DEX0291_147 DEX0134_89 DEX0291_272

DEX0291_148 flex DEX0134_89

DEX0291_149 DEX0134_90 DEX0291_273

DEX0291_150 flex DEX0134_90 DEX0291_274

DEX0291_151 DEX0134_91 DEX0291_275

DEX0291_152 flex DEX0134_91

DEX0291_153 DEX0134_92 DEX0291_276

DEX0291_154 flex DEX0134_92

DEX0291_155 DEX0134_93 DEX0291_277

DEX0291_156 flex DEX0134_93

DEX0291_157 DEX0134_94

DEX0291_158 DEX0134_95 DEX0291_278

DEX0291_159 DEX0134_96 DEX0291_279

DEX0291_160 DEX0134_97 DEX0291_280

DEX0291_161 flex DEX0134_97 DEX0291_281

DEX0291_162 DEX0134_98 DEX0291_282

DEX0291_163 DEX0134_99 DEX0291_283

DEX0291_164 flex DEX0134_99 DEX0291_284

The chromosomal locations were as follows:

DEX0291_6 chromosome 14

DEX0291_12	chromosome 2
DEX0291_16	chromosome 19
DEX0291_19	chromosome 17
DEX0291_21	chromosome 1
DEX0291_23	chromosome 12
DEX0291_24	chromosome 16
DEX0291_26	chromosome 10
DEX0291_28	chromosome 4
DEX0291_31	chromosome 11
DEX0291_33	chromosome 11
DEX0291_34	chromosome 11
DEX0291_36	chromosome 5
DEX0291_38	chromosome 11
DEX0291_39	chromosome 16
DEX0291_40	chromosome 13
DEX0291_41	chromosome 13
DEX0291_42	chromosome 10
DEX0291_43	chromosome 10
DEX0291_44	chromosome 7
DEX0291_46	chromosome 12
DEX0291_48	chromosome 10
DEX0291_52	chromosome 7
DEX0291_53	chromosome 7
DEX0291_56	chromosome X
DEX0291_58	chromosome 15
DEX0291_59	chromosome 7
DEX0291_60	chromosome 7
DEX0291_61	chromosome 5
DEX0291_62	chromosome 5

DEX0291_68	chromosome 5
DEX0291_70	chromosome 8
DEX0291_73	chromosome 2
DEX0291_77	chromosome 8
DEX0291_78	chromosome 16
DEX0291_79	chromosome 16
DEX0291_85	chromosome 15
DEX0291_89	chromosome 20
DEX0291_91	chromosome 2
DEX0291_93	chromosome 1
DEX0291_94	chromosome 15
DEX0291_98	chromosome 15
DEX0291_99	chromosome 2
DEX0291_101	chromosome 8
DEX0291_102	chromosome 8
DEX0291_107	chromosome 4
DEX0291_111	chromosome 7
DEX0291_118	chromosome 7
DEX0291_123	chromosome 4
DEX0291_124	chromosome 4
DEX0291_127	chromosome 6
DEX0291_128	chromosome 16
DEX0291_131	chromosome 3
DEX0291_132	chromosome 3
DEX0291_135	chromosome 1
DEX0291_138	chromosome 18
DEX0291_139	chromosome 18
DEX0291_141	chromosome 3
DEX0291_142	chromosome 1

DEX0291_146	chromosome 5
DEX0291_150	chromosome 1
DEX0291_151	chromosome 7
DEX0291_152	chromosome 7
DEX0291_153	chromosome 5
DEX0291_156	chromosome 12
DEX0291_160	chromosome 8
DEX0291_161	chromosome 8
DEX0291_162	chromosome 17
DEX0291_163	chromosome 6
DEX0291_164	chromosome 6

LSGs were also identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer" disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be

considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the LSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), highest differential expression for a given cancer (CLASP 4); differential expression in cancer tissue (CLASP 5), and. cDNA libraries were divided into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into "gene bins," where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with rigorous statistical significant testing taking into account variations in sample size and relative gene abundance in different libraries and within each library (for the equations used to determine statistically significant expression see Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were analyzed for statistical significance.

The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

(a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a gene must exhibit statistically significant expression in the tissue of interest compared to all other tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 1 candidate.

(b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and undetectable expression in libraries from normal individuals and libraries

from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.

(c) CLASP 4: highest differential expression for a given cancer: To qualify as a CLASP 4 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. In addition, it must be one of the 50 genes with the highest differential expression.

(d) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 5 candidate.

DEX0291_86 Lung 5 H

DEX0291_87 Lung 5 H

The individual tissue expression levels from the Incyte LifeSeq database were as follows:

DEX0291_1	SEQ ID NO: 1	UTR .0044	PAN .0059	BMR .0064	TON .0299
DEX0291_10	SEQ ID NO: 10	OVR .0021	PRO .0034	UNC .004	THR .0045
DEX0291_100	SEQ ID NO: 100	LMN .0028	NOS .0073		
DEX0291_101	SEQ ID NO: 101	INS .0067	GLB .0139	UTR .0182	NOS .022
DEX0291_104	SEQ ID NO: 104	UTR .0044	PAN .0059	BMR .0064	TON .0299
DEX0291_105	SEQ ID NO: 105	UTR .0044	PAN .0059	BMR .0064	TON .0299
DEX0291_108	SEQ ID NO: 108	LMN .0028	UNC .004		
DEX0291_11	SEQ ID NO: 11	FTS .0006	SPL .0021	LNG .0034	INS .0038
DEX0291_110	SEQ ID NO: 110	LIV .0038	THY .004	CRD .0068	
DEX0291_112	SEQ ID NO: 112	BLD .0225			
DEX0291_113	SEQ ID NO: 113	BLD .0225			
DEX0291_114	SEQ ID NO: 114	PNS .0047	CRD .0068	BON .0169	
DEX0291_115	SEQ ID NO: 115	PNS .0047	CRD .0068	BON .0169	
DEX0291_116	SEQ ID NO: 116	LNG .0006	OVR .001	PRO .0017	BLD .0048
DEX0291_117	SEQ ID NO: 117	LNG .0006	OVR .001	PRO .0017	BLD .0048
DEX0291_118	SEQ ID NO: 118	LNG .0006	LMN .0028		
DEX0291_119	SEQ ID NO: 119	PAN .0047	NOS .0073	GLB .0139	
DEX0291_12	SEQ ID NO: 12	FTS .0006	SPL .0021	LNG .0034	INS .0038
DEX0291_120	SEQ ID NO: 120	PAN .0047	NOS .0073	GLB .0139	
DEX0291_123	SEQ ID NO: 123	UNC .004	NOS .0073	LIV .0076	ADR .0089
DEX0291_124	SEQ ID NO: 124	UNC .004	NOS .0073	LIV .0076	ADR .0089
DEX0291_127	SEQ ID NO: 127	INS .001	BLD .0016	LNG .0017	MAM .0019
DEX0291_128	SEQ ID NO: 128	PNS .007			
DEX0291_13	SEQ ID NO: 13	OVR .0154	LNG .0296		
DEX0291_131	SEQ ID NO: 131	UTR .0006	LMN .0028	FAL .0063	
DEX0291_132	SEQ ID NO: 132	UTR .0006	LMN .0028	FAL .0063	
DEX0291_137	SEQ ID NO: 137	LMN .0056	PNS .0094	MAM .0194	FAL .0251
DEX0291_138	SEQ ID NO: 138	INS .001	MAM .0028	UNC .004	FAL .0126
DEX0291_139	SEQ ID NO: 139	INS .001	MAM .0028	UNC .004	FAL .0126

	DEX0291_14	SEQ ID NO: 14	OVR .0154	LNG .0296		
	DEX0291_142	SEQ ID NO: 142	THR .0023			
	DEX0291_146	SEQ ID NO: 146	PRO .0003	INL .0004	CON .0007	PAN .0008
	DEX0291_147	SEQ ID NO: 147	UTR .0075	PLE .0449		
5	DEX0291_148	SEQ ID NO: 148	UTR .0075	PLE .0449		
	DEX0291_15	SEQ ID NO: 15	INS .0789			
	DEX0291_151	SEQ ID NO: 151	UTR .0006	PAN .0012		
	DEX0291_152	SEQ ID NO: 152	UTR .0006	PAN .0012		
	DEX0291_155	SEQ ID NO: 155	BLV .0016	PRO .0017	MAM .0019	PNS .0023
10	DEX0291_158	SEQ ID NO: 158	OVR .0021	KID .0039	GLB .0046	FAL .0063
	DEX0291_160	SEQ ID NO: 160	THY .002			
	DEX0291_161	SEQ ID NO: 161	THY .002			
	DEX0291_18	SEQ ID NO: 18	BRN .0006	FAL .0063		
	DEX0291_19	SEQ ID NO: 19	UTR .0063	LMN .0167	BON .0225	
15	DEX0291_2	SEQ ID NO: 2	UTR .0044	PAN .0059	BMR .0064	TON .0299
	DEX0291_20	SEQ ID NO: 20	UTR .0063	LMN .0167	BON .0225	
	DEX0291_21	SEQ ID NO: 21	PAN .0353	LMN .0416	OVR .0503	INT .1052
	DEX0291_22	SEQ ID NO: 22	PAN .0353	LMN .0416	OVR .0503	INT .1052
	DEX0291_23	SEQ ID NO: 23	CRD .0114	KID .0128	ADR .0209	PLE .0449
20	DEX0291_25	SEQ ID NO: 25	LIV .0057			
	DEX0291_26	SEQ ID NO: 26	LIV .0057			
	DEX0291_27	SEQ ID NO: 27	PRO .0034	FAL .0063		
	DEX0291_29	SEQ ID NO: 29	UTR .0013	ADR .0015		
	DEX0291_33	SEQ ID NO: 33	MAM .0005	LNG .0006	ADR .0015	BLV .0016
25	DEX0291_34	SEQ ID NO: 34	MAM .0005	LNG .0006	ADR .0015	BLV .0016
	DEX0291_37	SEQ ID NO: 37	BRN .0031	THR .0045		
	DEX0291_38	SEQ ID NO: 38	BRN .0031	THR .0045		
	DEX0291_45	SEQ ID NO: 45	GLB .0093			
	DEX0291_47	SEQ ID NO: 47	PNS .0047	CRD .0068	BON .0169	
30	DEX0291_48	SEQ ID NO: 48	PNS .0047	CRD .0068	BON .0169	
	DEX0291_49	SEQ ID NO: 49	UTR .0263	NOS .066		
	DEX0291_5	SEQ ID NO: 5	THR .0091	BMR .0129	LMN .0139	
	DEX0291_51	SEQ ID NO: 51	LIV .0019	OVR .0031	URE .0112	
	DEX0291_53	SEQ ID NO: 53	PAN .0071	NOS .0073	LMN .0083	PRO .0119
35	DEX0291_54	SEQ ID NO: 54	PAN .0071	NOS .0073	LMN .0083	PRO .0119
	DEX0291_55	SEQ ID NO: 55	LNG .0006	OVR .001	PRO .0017	BLD .0048
	DEX0291_56	SEQ ID NO: 56	LNG .0006	OVR .001	PRO .0017	BLD .0048
	DEX0291_59	SEQ ID NO: 59	SPL .0042			
	DEX0291_65	SEQ ID NO: 65	FTS .0012	INS .0019	SPL .0021	KID .009
40	DEX0291_66	SEQ ID NO: 66	FTS .0012	INS .0019	SPL .0021	KID .009
	DEX0291_67	SEQ ID NO: 67	BRN .0023	LIV .0038	URE .0112	
	DEX0291_68	SEQ ID NO: 68	BRN .0023	LIV .0038	URE .0112	
	DEX0291_7	SEQ ID NO: 7	CON .0011			
	DEX0291_70	SEQ ID NO: 70	CRD .0023	BLD .0064		
45	DEX0291_71	SEQ ID NO: 71	PRO .0003	UTR .0004	BLO .0006	PRO .0006
	DEX0291_72	SEQ ID NO: 72	CON .0113	LIV .0189	ADR .0209	
	DEX0291_74	SEQ ID NO: 74	CON .0113	LIV .0189	ADR .0209	
	DEX0291_75	SEQ ID NO: 75	CON .0113	LIV .0189	ADR .0209	
	DEX0291_78	SEQ ID NO: 78	THY .006	KID .009	GLB .0093	LIV .0132
50	DEX0291_79	SEQ ID NO: 79	THY .006	KID .009	GLB .0093	LIV .0132
	DEX0291_8	SEQ ID NO: 8	CON .0011			
	DEX0291_80	SEQ ID NO: 80	NOS .0073	STO .0081	ESO .0102	
	DEX0291_86	SEQ ID NO: 86	BLO .0006	BLV .0006	INL .0012	LNG .0017
	DEX0291_87	SEQ ID NO: 87	BLO .0006	BLV .0006	INL .0012	LNG .0017
55	DEX0291_88	SEQ ID NO: 88	CRD .0023	PNS .0047	INT .015	URE .0225
	DEX0291_89	SEQ ID NO: 89	CRD .0023	PNS .0047	INT .015	URE .0225
	DEX0291_9	SEQ ID NO: 9	OVR .0021	PRO .0034	UNC .004	THR .0045
	DEX0291_90	SEQ ID NO: 90	LMN .0028	NOS .0073		

DEX0291_91	SEQ ID NO: 91	LMN .0028	NOS .0073	
DEX0291_92	SEQ ID NO: 92	INL .0045		
DEX0291_93	SEQ ID NO: 93	INL .0045		
DEX0291_95	SEQ ID NO: 95	CRD .0023	BLD .0064	
DEX0291_97	SEQ ID NO: 97	ADR .0104	KID .0141	PLE .015
DEX0291_98	SEQ ID NO: 98	ADR .0104	KID .0141	PLE .015
DEX0291_99	SEQ ID NO: 99	LMN .0028	NOS .0073	

Abbreviation for tissues:
BLO Blood; BRN Brain; CON Connective Tissue; CRD Heart; FTS Fetus; INL Intestine, Larg
INS Intestine, Small; KID Kidney; LIV Liver; LNG Lung; MAM Breast; MSL Muscles; NRV Nervous Tissue; OVR Ovary; PRO Prostate; STO Stomach; THR Thyroid Gland; TNS Tonsil Adenoids; UTR Uterus

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the LSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal tissue (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the LSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal tissue (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the LSNAs show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 164 being diagnostic markers for cancer.

Sequences	Sequence ID	ddx QPCR code
DEX0134_10	DEX0291_18	Lng261
DEX0134_17	DEX0291_29	Lng262
	DEX0291_30	
DEX0134_2	DEX0291_3	Lng259
	DEX0291_4	
DEX0134_24	DEX0291_40	Lng260
DEX0134_77	DEX0291_129	Lng264
	DEX0291_130	
DEX0134_80	DEX0291_134	Lng256
	DEX0291_135	
DEX0134_96	DEX0291_159	Lng228

DEX0134_10; DEX0291_18(SEQ ID NO:18); Lng261

Experiments are underway to test primers and probes for QPCR.

DEX0134_17; DEX0291_29(SEQ ID NO: 29); DEX0291_30(SEQ ID NO: 30); Lng262

Experiments are underway to test primers and probes for QPCR.

Primers Used for QPCR Expression Analysis in DEX291_29

Primer Probe Oligo	Start From	End To	queryLength	sbjctDescript
lng262For	33	50	18	DEX0134_17
lng262Rev	122	101	22	DEX0134_17
lng262Probe	54	74	21	DEX0134_17

DEX0134_2; DEX0291_3(SEQ ID NO: 3); DEX0291_4(SEQ ID NO:4) ; Lng259

Experiments are underway to test primers and probes for QPCR.

DEX0134_24; DEX0291_40(SEQ ID NO: 40) ; Lng260

Experiments are underway to test primers and probes for QPCR.

DEX0134_77; DEX0291_129(SEQ ID NO: 129); DEX0291_130(SEQ ID NO: 130); Lng264

Experiments are underway to test primers and probes for QPCR.

DEX0134_80; DEX0291_134(SEQ ID NO: 134); DEX0291_135(SEQ ID NO: 135); Lng256

Experiments are underway to test primers and probes for QPCR.

DEX0134_96; DEX0291_159(SEQ ID NO: 159); Lng228

Experiments are underway to test primers and probes for QPCR.

Primers Used for QPCR Expression Analysis

Primer Probe Oligo	Start From	End To	queryLength	sbjctDescript
lng228For	814	837	24	DEX0134_96
lng228Rev	940	917	24	DEX0134_96
lng228Probe	906	877	30	DEX0134_96

Example 3: Protein Expression

The LSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the LSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the LSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of LSNA, and six histidines, flanking the

COOH-terminus of the coding sequence of LSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of LSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column volumes of wash buffer. LSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine

serum (inactivated at about 56°C), and supplemented with about 10 g/1 of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, μg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

DEX0291_166 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
21-37	1.23	17

DEX0291_168 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
69-80	1.07	12

DEX0291_169 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
15-25	1.06	11

DEX0291_170 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
54-64	1.06	11

DEX0291_175 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
166-176	1.36	11
41-65	1.18	25

	190-207	1.15	18
	71-84	1.03	14
DEX0291_177	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	42-54	1.12	13
DEX0291_182	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	35-50	1.20	16
DEX0291_183	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	48-66	1.16	19
	68-87	1.13	20
DEX0291_184	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	52-93	1.05	42
DEX0291_185	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	199-209	1.24	11
	4-26	1.22	23
	322-353	1.09	32
	408-462	1.08	55
	467-482	1.01	16
	30-49	1.01	20
DEX0291_187	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	67-80	1.15	14
DEX0291_193	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	5-34	1.13	30
DEX0291_194	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	58-71	1.33	14
	195-210	1.06	16
	37-52	1.04	16
DEX0291_196	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	70-87	1.15	18
DEX0291_199	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	79-91	1.10	13
DEX0291_200	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	262-289	1.09	28
	225-234	1.07	10
	412-426	1.03	15
DEX0291_203	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	66-77	1.26	12
DEX0291_204	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	109-141	1.04	33
	61-78	1.03	18
	46-58	1.00	13
DEX0291_205	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	12-37	1.03	26
DEX0291_206	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length

	91-100	1.19	10
DEX0291_208	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	33-45	1.18	13
	105-122	1.10	18
	58-103	1.02	46
DEX0291_212	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	373-393	1.24	21
	70-81	1.20	12
	430-457	1.11	28
	485-533	1.07	49
	204-254	1.06	51
	289-314	1.04	26
	141-165	1.03	25
	462-478	1.03	17
	126-135	1.02	10
	172-202	1.02	31
	318-363	1.02	46
	10-34	1.01	25
	98-119	1.00	22
DEX0291_216	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	10-21	1.35	12
DEX0291_218	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	662-694	1.20	33
	36-61	1.12	26
	98-118	1.10	21
	283-334	1.02	52
	699-740	1.01	42
DEX0291_221	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	20-32	1.17	13
DEX0291_225	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	61-72	1.16	12
	3-58	1.07	56
DEX0291_226	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	7-20	1.02	14
DEX0291_228	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	73-83	1.05	11
	170-183	1.01	14
DEX0291_238	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	13-41	1.11	29
DEX0291_239	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	38-54	1.25	17
DEX0291_241	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	35-56	1.07	22
DEX0291_243	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	35-48	1.10	14
DEX0291_245	Antigenicity Index(Jameson-Wolf)		

positions	AI avg	length
144-155	1.04	12

DEX0291_247 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
44-57	1.14	14
93-107	1.06	15
69-84	1.02	16

DEX0291_249 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
262-289	1.09	28
225-234	1.07	10
412-426	1.03	15

DEX0291_251 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
91-100	1.19	10

DEX0291_255 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
14-25	1.18	12

DEX0291_256 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
12-21	1.11	10

DEX0291_257 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
21-31	1.19	11

DEX0291_259 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
595-607	1.25	13
446-457	1.15	12
80-92	1.09	13
632-641	1.08	10
246-257	1.06	12
1054-1073	1.06	20
336-383	1.05	48
955-975	1.02	21
1477-1505	1.02	29
425-439	1.01	15

DEX0291_264 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
22-32	1.02	11

DEX0291_276 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
53-73	1.06	21

DEX0291_280 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
32-48	1.04	17

DEX0291_281 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
34-48	1.23	15

DEX0291_282 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
58-113	1.10	56

DEX0291_284 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
111-131	1.05	21

The predicted helicities were as follows:

```
DEX0291_169 PredHel=2   Topology=o24-43i55-77o
DEX0291_170 PredHel=3   Topology=i29-48o63-82i94-116o
DEX0291_175 PredHel=5   Topology=i144-166o209-231i312-334o349-371i373-
395o
DEX0291_183 PredHel=1   Topology=o20-42i
DEX0291_190 PredHel=1   Topology=i57-79o
DEX0291_193 PredHel=1   Topology=o33-52i
DEX0291_195 PredHel=1   Topology=i21-38o
DEX0291_209 PredHel=1   Topology=o36-58i
DEX0291_213 PredHel=1   Topology=o20-37i
DEX0291_218 PredHel=1   Topology=o616-638i
DEX0291_223 PredHel=2   Topology=o20-42i55-86o
DEX0291_229 PredHel=2   Topology=i5-22o27-45i
DEX0291_239 PredHel=1   Topology=i58-80o
DEX0291_247 PredHel=2   Topology=o20-42i205-227o
DEX0291_254 PredHel=1   Topology=i7-29o
DEX0291_259 PredHel=6   Topology=o761-780i828-850o865-883i896-918o983-
1005i1035-1052o
DEX0291_260 PredHel=1   Topology=o55-77i
DEX0291_262 PredHel=1   Topology=o50-67i
DEX0291_270 PredHel=1   Topology=i20-39o
DEX0291_272 PredHel=1   Topology=o10-32i
DEX0291_279 PredHel=3   Topology=o42-64i99-121o126-148i
DEX0291_281 PredHel=1   Topology=i82-103o
DEX0291_283 PredHel=2   Topology=i13-35o55-77i
```

Examples of post-translational modifications (PTMs) of the LSP of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the LSPs of the invention (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html most recently accessed October 23, 2001).

DEX0291_165 Pkc_Phospho_Site 8-10;

DEX0291_166 Asn_Glycosylation 29-32; Ck2_Phospho_Site 9-12; Pkc_Phospho_Site 25-27;31-33;

DEX0291_167 Myristyl 4-9;30-35;32-37; Prokar_Lipoprotein 24-34;

DEX0291_169 Ck2_Phospho_Site 20-23; Pkc_Phospho_Site 20-22;90-92;

DEX0291_170 Ck2_Phospho_Site 59-62; Pkc_Phospho_Site 2-4;9-11;17-19;21-23;59-61;

DEX0291_172 Ck2_Phospho_Site 29-32;34-37; Myristyl 10-15;

DEX0291_173 Pkc_Phospho_Site 33-35;

DEX0291_175 Asn_Glycosylation 191-194;396-399; Ck2_Phospho_Site 164-167;308-311;344-347;405-408;414-417; Leucine_Zipper 291-312;298-319; Myristyl 97-102;174-179;176-181;258-263;375-380;431-436; Pkc_Phospho_Site 304-306;308-310; Tyr_Phospho_Site 62-70;

DEX0291_176 Camp_Phospho_Site 13-16; Ck2_Phospho_Site 5-8; Pkc_Phospho_Site 9-11;16-18;

DEX0291_177 Myristyl 20-25;52-57; Pkc_Phospho_Site 60-62;86-88;

DEX0291_178 Ck2_Phospho_Site 25-28;33-36;46-49; Pkc_Phospho_Site 15-17;50-52;

DEX0291_179 Ck2_Phospho_Site 58-61;80-83;84-87; Pkc_Phospho_Site 28-30;

DEX0291_181 Ck2_Phospho_Site 11-14;46-49; Pkc_Phospho_Site 2-4;46-48;

DEX0291_182 Ck2_Phospho_Site 16-19;30-33; Myristyl 20-25; Pkc_Phospho_Site 47-49;

DEX0291_183 Asn_Glycosylation 80-83; Ck2_Phospho_Site 82-85; Myristyl 26-31;30-35;63-68; Pkc_Phospho_Site 10-12;82-84;

DEX0291_184 Myristyl 16-21;40-45;44-49; Pkc_Phospho_Site 59-61;77-79;

DEX0291_185 Amidation 15-18;88-91; Asn_Glycosylation 186-189;396-399; Camp_Phospho_Site 476-479; Ck2_Phospho_Site 31-34;74-77;110-113;198-201;423-426; Fibrin_Ag_C_Domain 430-442; Myristyl 27-32;40-45;43-48;169-174;194-199;362-367;447-452;458-463;470-475; Pkc_Phospho_Site 336-338;392-394;411-413;474-476;479-481; Tyr_Phospho_Site 202-208;288-295;

DEX0291_186 Pkc_Phospho_Site 29-31;

DEX0291_187 Amidation 105-108; Bzip_Basic 125-139;126-139; Ck2_Phospho_Site 51-54;115-118; Glycosaminoglycan 23-26; Myristyl 2-7;37-42;96-101;102-107; Pkc_Phospho_Site 7-9;144-146;

DEX0291_188 Pkc_Phospho_Site 10-12;

DEX0291_189 Ck2_Phospho_Site 19-22; Pkc_Phospho_Site 3-5;41-43; Tyr_Phospho_Site 5-13;

DEX0291_190 Myristyl 49-54; Pkc_Phospho_Site 24-26; Prokar_Lipoprotein 38-48;

DEX0291_191 Ck2_Phospho_Site 35-38;95-98; Myristyl 6-11;

DEX0291_192 Myristyl 22-27; Pkc_Phospho_Site 43-45;

DEX0291_193 Camp_Phospho_Site 15-18; Myristyl 24-29;27-32; Pkc_Phospho_Site 47-49;55-57;

DEX0291_194 Asn_Glycosylation 30-33; Camp_Phospho_Site 62-65;120-123; Ck2_Phospho_Site 65-68;79-82;131-134;136-139;138-141;152-155; Myristyl 69-74;73-78;100-105;198-203; Pkc_Phospho_Site 79-81;133-135;193-195;

DEX0291_196 Asn_Glycosylation 32-35; Ck2_Phospho_Site 56-59;

DEX0291_197 Asn_Glycosylation 36-39; Myristyl 12-17; Pkc_Phospho_Site 20-22;

DEX0291_198 Ck2_Phospho_Site 73-76; Myristyl 12-17;17-22;66-71; Pkc_Phospho_Site 91-93;

DEX0291_199 Asn_Glycosylation 125-128; Camp_Phospho_Site 70-73; Ck2_Phospho_Site 118-121; Myristyl 61-66;131-136; Pkc_Phospho_Site 32-34;68-70;80-82;89-91;140-142;153-155;

DEX0291_200 Amidation 2-5; Asn_Glycosylation 74-77;127-130; Ck2_Phospho_Site 76-79;230-233;242-245;262-265;423-426;566-569; Leucine_Zipper 95-116; Myristyl 34-39;419-424;499-504;536-541; Pkc_Phospho_Site 276-278;380-382;387-389;442-444;591-593; Tyr_Phospho_Site 561-568;562-568;

DEX0291_202 Myristyl 5-10;

DEX0291_203 Ck2_Phospho_Site 31-34;48-51; Myristyl 4-9; Pkc_Phospho_Site 84-86; Prokar_Lipoprotein 39-49;

DEX0291_204	Asn_Glycosylation 64-67; Ck2_Phospho_Site 33-36;166-169;188-191;234-237;236-239; Pkc_Phospho_Site 66-68;140-142;243-245;249-251; Tyr_Phospho_Site 124-132;
DEX0291_205	Camp_Phospho_Site 22-25; Myristyl 70-75; Pkc_Phospho_Site 13-15;
DEX0291_206	Ck2_Phospho_Site 66-69;96-99; Glycosaminoglycan 50-53; Myristyl 47-52;49-54;53-58;62-67; Pkc_Phospho_Site 12-14;132-134;141-143;192-194;210-212; Prokar_Lipoprotein 159-169;
DEX0291_207	Asn_Glycosylation 70-73; Ck2_Phospho_Site 15-18;81-84;
DEX0291_208	Asn_Glycosylation 75-78; Camp_Phospho_Site 84-87; Ck2_Phospho_Site 36-39;78-81; Pkc_Phospho_Site 111-113;
DEX0291_209	Asn_Glycosylation 30-33; Myristyl 5-10; Pkc_Phospho_Site 26-28;
DEX0291_210	Asn_Glycosylation 56-59;90-93;173-176; Ck2_Phospho_Site 24-27;58-61;92-95;138-141;148-151;270-273; Myristyl 45-50;49-54;306-311;312-317;354-359;397-402;426-431;433-438; Peroxidase_2 252-263; Pkc_Phospho_Site 70-72;138-140;145-147;163-165;195-197;207-209;252-254;263-265;275-277;325-327;359-361;402-404;444-446; Wd_Repeats 267-281;351-365;
DEX0291_211	Myristyl 5-10;
DEX0291_212	Amidation 102-105;521-524; Asn_Glycosylation 33-36;62-65;201-204;230-233;313-316;342-345;343-346;454-457;536-539; Ck2_Phospho_Site 194-197;204-207;388-391;511-514; Myristyl 183-188;220-225; Pkc_Phospho_Site 6-8;143-145;177-179;286-288;345-347;398-400;404-406;521-523; Zinc_Finger_C2h2 25-45;53-73;81-101;109-129;137-157;165-185;193-213;221-241;249-269;277-297;305-325;333-353;361-381;389-409;417-437;445-465;473-493;
DEX0291_213	Myristyl 24-29;
DEX0291_214	Pkc_Phospho_Site 9-11;
DEX0291_215	Myristyl 5-10;8-13;11-16;
DEX0291_216	Camp_Phospho_Site 18-21; Pkc_Phospho_Site 17-19;60-62;
DEX0291_217	Pkc_Phospho_Site 30-32;
DEX0291_218	Asn_Glycosylation 72-75;261-264;370-373;474-477;516-519; Camp_Phospho_Site 224-227;366-369; Ck2_Phospho_Site 36-39;180-183;253-256;333-336;380-383;457-460;778-781; Myristyl 177-182;217-222;266-271;319-324;368-373;381-386;384-389;393-398;482-487;575-580;585-590;649-654;731-736;732-737; Pkc_Phospho_Site 50-52;151-153;315-317;475-477;507-509;513-515;637-639;653-655;694-696; Tyr_Phospho_Site 193-200;290-296;681-688;
DEX0291_219	Ck2_Phospho_Site 39-42; Myristyl 47-52;48-53; Pkc_Phospho_Site 39-41;
DEX0291_220	Asn_Glycosylation 20-23; Ck2_Phospho_Site 123-126; Glycosaminoglycan 72-75; Myristyl 30-35;75-80; Pkc_Phospho_Site 123-125; Prokar_Lipoprotein 70-80; Tyr_Phospho_Site 107-114;
DEX0291_221	Amidation 29-32; Asn_Glycosylation 23-26; Ck2_Phospho_Site 36-39; Pkc_Phospho_Site 24-26;36-38;

DEX0291_222 Ck2_Phospho_Site 78-81; Myristyl 34-39;36-41; Pkc_Phospho_Site 96-98;

DEX0291_223 Myristyl 27-32;31-36;33-38;60-65;64-69;66-71;67-72;70-75;77-82;84-89;93-98;95-100;98-103; Prokar_Lipoprotein 27-37;58-68;

DEX0291_224 Asn_Glycosylation 30-33;181-184; Camp_Phospho_Site 37-40; Ck2_Phospho_Site 7-10; Pkc_Phospho_Site 83-85;

DEX0291_225 Asn_Glycosylation 24-27; Ck2_Phospho_Site 3-6;72-75; Myristyl 20-25; Pkc_Phospho_Site 48-50;

DEX0291_226 Pkc_Phospho_Site 19-21;

DEX0291_227 Leucine_Zipper 10-31; Pkc_Phospho_Site 3-5;

DEX0291_228 Asn_Glycosylation 182-185; Camp_Phospho_Site 27-30; Ck2_Phospho_Site 21-24;41-44;78-81;98-101;112-115; Myristyl 53-58;96-101;103-108; Pkc_Phospho_Site 9-11;58-60; Tyr_Phospho_Site 76-82;

DEX0291_229 Myristyl 13-18;

DEX0291_231 Pkc_Phospho_Site 16-18;

DEX0291_233 Ck2_Phospho_Site 79-82; Myristyl 22-27;47-52; Pkc_Phospho_Site 15-17;97-99; Prokar_Lipoprotein 42-52;

DEX0291_235 Asn_Glycosylation 44-47; Myristyl 30-35;

DEX0291_236 Ck2_Phospho_Site 40-43; Pkc_Phospho_Site 40-42;

DEX0291_237 Ck2_Phospho_Site 13-16; Myristyl 19-24;23-28;54-59; Pkc_Phospho_Site 32-34;95-97;

DEX0291_238 Amidation 14-17; Pkc_Phospho_Site 14-16;

DEX0291_239 Ck2_Phospho_Site 30-33; Myristyl 52-57;53-58;71-76; Pkc_Phospho_Site 49-51;

DEX0291_240 Amidation 119-122; Camp_Phospho_Site 121-124; Ck2_Phospho_Site 124-127;181-184; Glycosaminoglycan 115-118; Myristyl 5-10;13-18;14-19;16-21;18-23;19-24;20-25;21-26;22-27;23-28;33-38;36-41;37-42;55-60;90-95;103-108;105-110;106-111;108-113;116-121;136-141;170-175; Pkc_Phospho_Site 9-11;

DEX0291_241 Asn_Glycosylation 19-22;53-56; Ck2_Phospho_Site 4-7;80-83; Myristyl 52-57; Pkc_Phospho_Site 103-105;

DEX0291_242 Leucine_Zipper 10-31;

DEX0291_243 Myristyl 10-15;57-62;

DEX0291_244 Rgd 15-17;

DEX0291_245 Ck2_Phospho_Site 144-147; Myristyl 5-10;165-170;

DEX0291_247 Asn_Glycosylation 73-76;101-104;167-170; Camp_Phospho_Site 230-233; Ck2_Phospho_Site 159-162;194-197; Glycosaminoglycan 33-36; Leucine_Zipper 198-219; Myristyl 2-7;34-39;74-79;87-92;112-117;116-121;119-124;149-154;164-169;186-191;217-222; Pkc_Phospho_Site 43-45;77-79;129-131;134-136;171-173;

DEX0291_248 Ck2_Phospho_Site 34-37; Myristyl 43-48; Pkc_Phospho_Site 58-60; Tyr_Phospho_Site 16-24;

DEX0291_249 Amidation 2-5; Asn_Glycosylation 74-77;127-130; Ck2_Phospho_Site 76-79;230-233;242-245;262-265;423-426;566-569; Leucine_Zipper 95-116; Myristyl 34-39;419-

	424;499-504;536-541; Pkc_Phospho_Site 276-278;380-382;387-389;442-444;591-593; Tyr_Phospho_Site 561-568;562-568;
DEX0291_251	Ck2_Phospho_Site 66-69;96-99; Glycosaminoglycan 50-53; Myristyl 47-52;49-54;53-58;62-67; Pkc_Phospho_Site 12-14;132-134;141-143;192-194;210-212; Prokar_Lipoprotein 159-169;
DEX0291_252	Camp_Phospho_Site 27-30; Myristyl 15-20; Pkc_Phospho_Site 5-7;
DEX0291_253	Asn_Glycosylation 55-58;
DEX0291_254	Pkc_Phospho_Site 9-11;
DEX0291_255	Camp_Phospho_Site 20-23; Myristyl 67-72;
DEX0291_256	Camp_Phospho_Site 15-18; Ck2_Phospho_Site 23-26; Pkc_Phospho_Site 18-20;30-32;
DEX0291_257	Asn_Glycosylation 2-5;27-30; Myristyl 15-20; Pkc_Phospho_Site 4-6;
DEX0291_258	Ck2_Phospho_Site 53-56;
DEX0291_259	Amidation 362-365;513-516;968-971; Asn_Glycosylation 133-136;144-147;233-236;298-301;478-481;601-604;635-638;638-641;830-833; Ck2_Phospho_Site 9-12;235-238;300-303;343-346;459-462;587-590;698-701;706-709;788-791; Myristyl 35-40;53-58;68-73;69-74;102-107;211-216;229-234;296-301;473-478;728-733;747-752; Pkc_Phospho_Site 86-88;212-214;235-237;343-345;353-355;480-482;617-619;706-708;729-731;818-820;925-927; Prokar_Lipoprotein 978-988; Tyr_Phospho_Site 697-704;891-898;
DEX0291_260	Ck2_Phospho_Site 17-20;49-52;77-80; Pkc_Phospho_Site 45-47;
DEX0291_261	Pkc_Phospho_Site 5-7;32-34;
DEX0291_262	Asn_Glycosylation 76-79; Ck2_Phospho_Site 16-19;45-48; Myristyl 6-11;9-14;56-61;58-63; Pkc_Phospho_Site 25-27;84-86;
DEX0291_264	Asn_Glycosylation 19-22; Camp_Phospho_Site 31-34; Ck2_Phospho_Site 40-43;74-77; Myristyl 37-42;90-95;
DEX0291_267	Asn_Glycosylation 56-59;98-101; Myristyl 66-71;
DEX0291_268	Pkc_Phospho_Site 33-35;
DEX0291_269	Ck2_Phospho_Site 8-11;
DEX0291_270	Asn_Glycosylation 3-6; Ck2_Phospho_Site 6-9;
DEX0291_271	Myristyl 10-15;
DEX0291_272	Myristyl 9-14; Pkc_Phospho_Site 3-5;
DEX0291_273	Ck2_Phospho_Site 29-32; Pkc_Phospho_Site 29-31;
DEX0291_274	Ck2_Phospho_Site 215-218; Pkc_Phospho_Site 184-186;
DEX0291_275	Myristyl 2-7; Pkc_Phospho_Site 15-17;
DEX0291_276	Ck2_Phospho_Site 5-8; Pkc_Phospho_Site 21-23;33-35;44-46;65-67;
DEX0291_279	Myristyl 78-83;122-127; Pkc_Phospho_Site 25-27; Prokar_Lipoprotein 49-59;
DEX0291_280	Asn_Glycosylation 11-14; Camp_Phospho_Site 33-36;34-37; Pkc_Phospho_Site 32-34;37-39;51-53;
DEX0291_281	Ck2_Phospho_Site 35-38;44-47;

DEX0291_282 Camp_Phospho_Site 109-112;123-126;161-164; Ck2_Phospho_Site 98-101;141-144;150-153;151-154; Myristyl 38-43; Pkc_Phospho_Site 23-25;71-73;78-80;98-100;133-135;141-143;150-152;155-157;156-158;

DEX0291_283 Asn_Glycosylation 2-5; Myristyl 19-24; Pkc_Phospho_Site 51-53;

DEX0291_284 Ck2_Phospho_Site 5-8;14-17;25-28;102-105;137-140;148-151; Myristyl 60-65;91-96;

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 164. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al., Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al., Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al., Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al., Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and

chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 μg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of

administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad.

Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 μg/kg of the polypeptide for six consecutive days. Preferably, the

polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 1. Preferably, the 5'primer contains an EcoRI site and the 3'primer

includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known

in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by

the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about

0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon,"Transgenic Animals," Intl.

Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR

(rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such

approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the

development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 165 through 284;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 164;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 60% sequence identity to the nucleic acid molecule of (a) or (b).

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.

5. The nucleic acid molecule according to claim 4, wherein the nucleic acid molecule is a human nucleic acid molecule.

6. A method for determining the presence of a lung specific nucleic acid (LSNA) in a sample, comprising the steps of:

(a) contacting the sample with the nucleic acid molecule according to claim 1 under conditions in which the nucleic acid molecule will selectively hybridize to a lung specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to a LSNA in the sample, wherein the detection of the hybridization indicates the presence of a LSNA in the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 165 through 284 ; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 164.

12. An antibody or fragment thereof that specifically binds to the polypeptide according to claim 11.

13. A method for determining the presence of a lung specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the lung specific protein; and

(b) detecting binding of the antibody to a lung specific protein in the sample, wherein the detection of binding indicates the presence of a lung specific protein in the sample.

14. A method for diagnosing and monitoring the presence and metastases of lung cancer in a patient, comprising the steps of:

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the lung specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of lung cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with lung cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the lung cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

SEQUENCE LISTING

<110> Macina, Roberto
Recipon, Herve
Chen, Sei-Yu
Sun, Yongming
Liu, Chenghua
diaDexus, Inc.

<120> Compositions and Methods Relating to Lung Specific Genes and Proteins

<130> DEX-0292

<150> 60/252,500
<151> 2000-11-22

<160> 284

<170> PatentIn version 3.1

<210> 1
<211> 338
<212> DNA
<213> Homo sapien

<400> 1
tggatacgag cggacagccc gggcaggtac cgcatactag caaaggtaat ggtgatctag 60

caaacaaaat tggtttctag cagttagaag tgagcaggag cacttgtatt atagtattta 120

aataatcctg gttaatctct ttttaagccg agtaacccct ccagattttg cctttttatt 180

attgaggctg gctttatttt cttctacttt ttttcccgtt ttatagcagt taattatttt 240

tgtgattatt atgcaagaag cattgccctt gagttaaact gttattgttt cataagcagc 300

tattaaaata actgagcatg ttttatgaac atacacta 338

<210> 2
<211> 2446
<212> DNA
<213> Homo sapien

<400> 2
aaggatcctt aattaaatta atccccccccc ccctttttttt ttgtattctt gccagtacag 60

tatatggttt ttctacccca attacatact gggttttgta ccacatcact aaaggcccaa 120

atcattgaag atacaaaacc gtacatgcag gctggttgtc tggttagtca atggctgatt 180

tgcttcaact gtctagtatg tatgtgcagc ctgaaactgg ctccttaaaa ggaaagccgg 240

gtcagtcatc ttgaaaaaat gacatgtaaa agtaaatcga taattgtttt gagagacggt 300

acatgtttta aaggttggcc ttaagcttca gtaacattgt cattttgtga ccttttgttg 360

tcacacctgt accctaacct gacaggaatt aactactgtt tttttgtggg gcagaaagca 420

aaacctggtg ttgtgacttt tatcctaatg gttcttaggc aaggttagtg agaagaaaca 480

caaacccaga tgcatgcatt gtgcattatt ttgtagacaa gctacttttt cttctgtccc 540

tttaacaaat ttgcagcaat taccctccct ttggggtcta gagtgaaagc taatttgtgg 600

gtagatgaga ttgcagaaga atggatgtcc atggctgtga acactgcaca ctgcacatcc 660

atctccagtg ctcacactgt gcagctacca ctccctggct gcgtgccatg ctgtcgggtt 720

gcagatttgc acacataaat tcctcaggaa gagtttgcat gagcatcacc tcgcaatatt 780

ctgtactgac caaacaaggg atttgaacgt ttttcagcac aaaaggataa cttccgagtg 840

gtggtctgta cgcatactag caaaggtaat ggtgatctag caaacaaaat tggtttctgc 900

agttagaagt gagcaggagc acttgtatta tagtatttaa ataatcctgg ttaatctctt 960

tttaagccga gtaacccctc cagattttgc ctttttatta ttgaggctgg ctttattttc 1020

ttctactttt tttcccgttt tatagcagtt aattattttt gtgattatta tgcaagaagc 1080

attgcccttg agttaaactg ttattgtttc ataagcagct attaaaataa ctgagcattg 1140

ttttatgaac atacactaat ctgagatact gaaaagcttt gcaactaaaa agcaaaacaa 1200

cctacattag tcatctagcc attgtttgga tgttttgagt tgatttttta tggtgcctct 1260

tttagcttgg aatattacgt ttactttaat ccaagtctag gccttttaaa gggtccttaa 1320

aattaaagtt cagaatgtga atccctttga catctattac aggtttatag gacctttttg 1380

gttgtgatta ctgttttcaa tacgattgta taaatgaagt taactttgtc agaagttaaa 1440

atggaggtca taggagttcc tggagaaatg gctctcctgt ttctttcatt accccactga 1500

agttcacccc agtttctggc cacaagaata tgagaaagga accctgttgt tttccaaggg 1560

aaatcattcc tctctgtccc cactgttgat taactaaagt cctggacacc ttccttcctc 1620

cactggccaa gacccacctt gacccacctt gaacctcttt tcagagccga gtggcatgaa 1680

tatgtgtact gtttctgctt ctgttgatgg agtggctgtg ggagaattaa aggaaatgct 1740

aatttgagct tcattcatag gggaacctac tatatattgc atccctgctg gttggaaatt 1800

atcttcatct ctggactgca ttgtttagaa aaatgttaat ggcttacaat tctgagaact 1860

ttattgtgtg gctctggggt taagaattct gtggtttgaa aaaaaataaa tattttgtat 1920

tgattctcac gtcatttcaa tgttgtgact atgtactaaa tgcactaaga ctgggtattc 1980

tcttagaaga gtgcgttttg ttaaacagat ggcagttcac tctcattagt tctatttgtc 2040

aatattacag ccacacactt caggataact tactcaaatg tgaagtcatg ggaagctatg 2100

gctagtaaca ggaatgcttg tgaaaaattt ggagggcaag cacgagcaga ggcttttgcc 2160

actcactagc acagccagag caaggatgaa gcatgaagtc ttgttcctag tggtgcttgt 2220

atgtcagaag ccatagtgag ctcagcccgg gcgttctgca catcatctct ttaatcccag 2280

aaaaacaaat ggcggatgaa actcctattg acctgggata caagccctga ctttggctct 2340

tgatcttttg gtgggttctg aatcactcct tctgtattag ggcatatact ttcactctaa 2400

aacttgtggt aaatgaaata aatcttgtac agatgtaaaa aaaaaa 2446

<210> 3
<211> 460
<212> DNA
<213> Homo sapien

<400> 3
acacacccaa atgtattcca agaccatact ctgtgacagc tgaaagatca aatttttgc 60

ttgctatcca gagatgagcc ttctgctcca ctgggaactt ggctctacca gctcacagac 120

ccctttggga agctgcagaa ataaaacact cccctctatg cctcactaca caagattaat 180

acacttctag tgaaagaaga attttaatgc acttgagatg aaagaagcca gatgctttcc 240

tatttatgtc ctttccaacc atttagagtg cgggggatgg aagtgtagca aggcttgacc 300

tgcaactccc aggaccctgc tggtcatgtc aatcccaagg gcagaaatat ctcttctaga 360

atccttccag ctcacctcca ctgtggctac aagtgaaagc cataagtcaa acgggagctg 420

caggaagcca cacttactgc actgtccaag aataaatcaa 460

<210> 4
<211> 594
<212> DNA
<213> Homo sapien

<400> 4
acacacccaa atgtattcca agaccatact ctgtgacagc tgaaagatca aatttttgc 60

ttgctatcca gagatgagcc ttctgctcca ctgggaactt ggctctacca gctcacagac 120

ccctttggga agctgcagaa ataaaacact cccctctatg cctcactaca caagattaat 180

acacttctag tgaaagaaga attttaatgc acttgagatg aaagaagcca gatgctttcc 240

tatttatgtc ctttccaacc atttagagtg cgggggatgg aagtgtagca aggcttgacc 300

tgcaactccc aggaccctgc tggtcatgtc aatcccaagg gcagaaatat ctcttctaga 360

atccttccag ctcacctcca ctgtggctac aagtgaaagc cataagtcaa acgggagctg 420

caggaagcca cacttactgc actgtccaag aataaatcaa aagcttttaa tagcaagaag 480

ttaaaaataa agcagcacgt ccttaaaaat aaaatttcct acaactgaca aaataaagga 540

cgaacacgca ctggttgttt cttaaccact agtggacaag caaagggaaa aagc 594

<210> 5
<211> 283
<212> DNA
<213> Homo sapien

<400> 5
actactaaag gcaaccggca tggactcata atatttgtgg ggacagcaaa aagactaaaa 60

gttctaagga agaaaatgcg aaccttgata gtttgaaata gttaaaaaga cagtgtagaa 120

actgcttagg cagttggatt atggactatt agatgatact tgggtctgat aatggtataa 180

ggagaataaa gtatttaggg atccaatatt acgcctgcag ctttttccaa atagttcatg 240

ggggaggggg atgtgtaagt ggttaactga agtctaacta gat 283

<210> 6
<211> 2237
<212> DNA
<213> Homo sapien

<400> 6
ttttttttttt caagaaaaaa aaatcacttt aattgaggaa cactttcagt ttgtgacaaa 60

attatgctgt gaatcaggtg ttgcaaatta tggcccactg cctgcttttg tgtaagtttt 120

attggaacac agctacattc agtccatggc tgcttttaga atacaacagt agactttaac 180

atttggaaca gggaacagaa accagagcca tacagctaat aaacttgaaa atatttacaa 240

gttgatgctt tacaaaatcc atctgctgac ccctgctctg taccattgtt ctcttctgat 300

ggtctgttta ctaaaaaata aaaacttcac aaacatgtaa aaaatagatt tgccatttaa 360

aatgtgcttt tcaagtttga ctttttagga tgcaattaat tcactaaata cagaacttaa 420

ctaaggacaa aatttaaaga tcagcattct ttcccttccc atcacgctca acttaacatg 480

aagaactgta aacatcctaa gcttacaaca aacctatcta gttagacttc agttaaccac 540

ttacacatcc ccctccccca tgaactattt ggaaaaagct gcaggcgtaa tattggatcc 600

ctaaatactt tattctcctt ataccattat cagacccaag tatcatctaa tagtccataa 660

tcaaactgcc taagcagttt ctacactgtc tttttaacta tttcaaacta tcaaggttcg 720

cattttcttc cttagaactt ttagtctttt tcttccccaa aatatttgag tccatgccag 780

ttgcctttag ttgtacccaa ataatggttt gtctatttcc taaaagtagt actcttaaat 840

ttaaatttag tgttattttt gttgtcattg ttccttcttc ctcatgtggt tgtgcaggca 900

gagcttgagc atccagattt caaaattaaa aaataaaaga taatctagtt taatatatag 960

tagttgaatc accttaagtc tagactgctg tatgagcacc cattatcttt cactatattc 1020

catcatcccc caacatatcc acagtagatg aagggcagtt tgctcaaaca ttgttttgat 1080

cctgtcatgt ctgttcagaa atgcctgtct attcagaaac ccacgtctaa taacaaaatc 1140

ttggactggt tactatcaaa acccaacaac atacagactc ctcagctagg ccctagggat 1200

atttttctac cttgatttcc aaatgttcat tgaaagaatg cttaattcta atttggaaaa 1260

aagtttttgg cttcccactt ctgctttaca cgttcatctt tcttgaaatc aaatccaatc 1320

caatctatat tctaagaacc tgctcaaatc ttggttcttc aaagctttcc ctggtatttt 1380

gcatttttgc tttgaatagt tccacgaagg ataacttctt actccttcct tcatctttct 1440

gtatcttgca tatagtaaat attaatgact tgtttgcatt ttgttatcct aacttggcta 1500

taaagaaaat cagatgtctt caccagtcgt tcaaacttca ggtctgccta cagattcata 1560

gatggctgtg gatttttata attttgtcac aaagttagtg gtaactacag gttatctcag 1620

aatatctttt ttggcgtata aatttttttc ttttcctttt ttagacagtc tctctctctg 1680

tcgcccaggc tagagtgcag tggcgtgatc ccggctcact acaacctctg cctcctgggt 1740

tcaagagatt cttagcctca gcctcccgag tagctagggt tacaggcgcg caccacctcc 1800

atgcccagct cttttgtatt tttaagtaga gacagggttt caccatgttg gtcaggctgg 1860

tctcgaactt ctgacttcag gcaaccttag ccgcctcggc ctcccaaagt gctgggatta 1920

caggcacaag ccactgcacc cagccttatt accataaatc atcttgatgc tggtacctga 1980

taagattcta tttgcttttc tttattcata gagaccacaa acagatcgca gatccaggtt 2040

tctcaaactg gagcatctgc ttaatttttcc cataaaatca gtcttattct ttctgacagc 2100

tctgagactc ctccggccac gactaggtgc tgtcctggag gaaacggtgg aggacggccg 2160

cacaaaaacc aatctacctg atgaaaactc cgttcccttc tcgccagaaa cataaaatgc 2220

gatggatacg ctcgtgc 2237

<210> 7
<211> 777
<212> DNA
<213> Homo sapien

<400> 7
tggggtcgcg gcgaggtacc tttctgtttt aaaatgttca aataatgtat tttttaaaaa 60

tgtgtaatca ttcaaggaaa cgggaaaaca agtctaaata aaaccgggtc ctaaagttta 120

ctctaggagt cattattctt cctttgcagt ctcaattcta tttactccgt aacaagtaaa 180

ctgttttatc ccttcggctg ggaaggataa tttttaccct accaagatga tgacaatgcc 240

actgcagtgc tcacccagca agaaacacat aaatatgaag cagcgtcaaa gtctttctat 300

gaagtgtatg caagaattga tttgcaatgc attaattcag gaattcagtg gagaggctcc 360

tactatacct agcagatcta aacttgagct catgataaca ctgcacaaag ctatggtaga 420

taaaagtgat aggtaatgcc aatagaacaa ttccactgac aacacaaact cctccaagaa 480

ttcttccagg cactgtgata ggatacatat ctccatagcc aactgtagtc atagagataa 540

tcacccacca gcaggcagca ggaatgctgg taaagtcctt gttggatgtt tccaggtcca 600

gcccatgttc aagaagctga gaaagtgcac taaagattgc catggcaaca caaatgaaga 660

cagtaacata accatctctc ggtagcacgt ttgagagtca accgagtgct gaagaccatt 720

aagtgacggc agctggcgta atcatgtcta gctgttctgt gtgaatgttc gtcaacc 777

<210> 8
<211> 911
<212> DNA
<213> Homo sapien

<400> 8
ggttgacgaa cattcacaca gaacagctag acatgattac gccagctgcc gtcacttaat 60

ggtcttcagc actcggttga ctctcaaacg tgctaccgag agatggttat gttactgtct 120

tcatttgtgt tgccatggca atctttagtg cactttctca gcttcttgaa catgggctgg 180

acctggaaac atccaacaag gactttacca gcattcctgc tgcctgctgt gggtgattat 240

ctctatgact acagttggct atggagatat gtatcctatc acagtgcctg gaagaattct 300

tggaggagtt tgtgttgtca gtggaattgt tctattggca ttacctatca cttttatcta 360

ccatagcttt gtgcagtgtt atcatgagct caagtttaga tctgctaggt atagtaggag 420

cctctccact gaattcctga attaatgcat tgcaaatcaa ttcttgcata cacttcatag 480

aaagactttg atgctgcttc atatttatgt gtttcttgct gggtgagcac tgcagtggca 540

ttgtcatcat cttggtaggg taaaaattat ccttcccagc cgaagggata aaacagttta 600

cttgttatgg agtaaataga attgagactg caaaggaaga ataatgactc ctagagtaaa 660

ctttaggacc cggttttatt tagacttgtt ttcccgtttc cttgaatgat tacacatttt 720

taaaaaatac attatttgaa cattttaaaa cagaaaggta ctattttcca atgtttttcc 780

atcttatgaa ttcagaagaa gcttggaact tatagtgttt tttgtttgag agtaacattt 840

tcatttctaa atgttttata atttctcata tcaatgtcag aagtatcctg gaaacatatg 900

tcacatgcga g 911

<210> 9
<211> 445
<212> DNA

<213> Homo sapien

<400> 9
gccgcccggg caggtacatg tgcacttaaa tgtaatagca ccaacattca ttatattatg 60

aaggggatac tttataagaa ttataaatta tttttacatg attaaataat tttggcagag 120

taagtccgca ggacttaaat aaccagtcag ccttagtatc tacatctgga ccaggaagcc 180

ttgtgttaca gacacagatt ccctgtgaag ttctccaggg tgtaaagaag gccccaggga 240

ggtgccgacg ggatgtgagg gctgcctttc agctcaggcg agcctgcaag caacagggca 300

gttgggaagg gtcacaggca caggaaggga gcagtgggca accattgatt acatcaaacc 360

aggatggttt cttcttttta aagaaacttt tgtgagtgtt tacccacccc catccacaca 420

aatatgctct ctaggaaatg tgaaa 445

<210> 10
<211> 1254
<212> DNA
<213> Homo sapien

<400> 10
gtccgcttaa ttaaagatct ttttttttta ttttttttta ttgttaagca tatttgtata 60

ttttttacta gttatttcat acttgccctg aaagaataca cattcaaaaa gcttgaaatt 120

aggcaatgtc agtctcatca aacaaaacca gcattggaag cgaatattaa caaatatcag 180

aatgaaatta caaaatatac atctccagcc tcataaaaca tgaattttat aagcactttc 240

ataaacataa gaaaaatagc tttgacaata actatttgaa aacattatta ttaataatct 300

tgtatactgt acatgtgcac ttaaatgtaa tagcaccaac attcattata ttatgaaggg 360

gatactttat aagaattata aattattttt acatgattaa ataattttgg cagagtaagt 420

ccgcaggact taaataacca gtcagcctta gtatctacat ctggaccagg aagcctttgt 480

gttacagaca cagattccct gtgaagttct ccagggtgta aagaaggccc cagggaggtg 540

ccgacgggat gtgagggctg cctttcagct caggcgagcc tgcaagcaac agggcagttt 600

gggaagggtc acaggcacag gaagggagca gtgggcaacc attgattaca tcaaaccagg 660

atggtttctt ctttttaaag aaacttttgt tgagtgttta cccacccca tccacacaaa 720

tatgctctct aggaaatgtg aaagtataag cttcagaaaa atgtttttct catcctttaa 780

tttctggttt tatgatcaaa tgttccgaag acactgcttc tttttgctca aaatagttga 840

cattgtcatt gcttctagtt tttcagctcc tactatgctt actatgtcca tgggtgccat 900

tgacgtcttt ggctatttat catctctaga aaagaaaaaa aaaacattaa aattttagga 960

tgtaggattc acaataaatc tcttaccaaa tcctccttaa atccctaaac tctctacctg 1020

gctctgactc caggatcacc ctttcagttc tttccatgtt tttcttcgga aaatgtagtt 1080

tcctaacctt tctacccacc cttctcgtct actcttaggt ctgcagccac agtgatagtt 1140

ctaagttgca aacctgatca tgtccttcca gtacttagtc gttgaaaaat cttcctactt 1200

cccgcagatt gaagcctgaa ctctttaata ggcatctaag gacttggacc cgga 1254

<210> 11
<211> 838
<212> DNA
<213> Homo sapien

<400> 11
gcccgggcag gtacttaact aattgggctg aggatgaata tatcagccac agcacattaa 60

agaatgagcc aaggattgtc atggttggtc acttttaaa gtattgatta ctgcaactgg 120

agaatggaaa gtgtatattg gtgacgccaa cctcagtttc tgagcactcc tgctctgtgg 180

tgagaatcag acaaaaattc atcggggtga aaaaggcatt acctgattca caccttgtc 240

ttgctagccc tcttccattc atttctcaca cagcacttgc tctgttaaat cctctctctg 300

tctcagacca tgcttgcccc ttcaaagggt atggttcagg ctcctttcaa gacattggag 360

tttctctctg gggaaagaga gccccctact ggtttggctt cagtctaggt ccaccatccc 420

tctcgatctg gcatcttgga gattaattta aaaggcaagc tcaccacaat gtaagcctat 480

ggtctggcca accttgcttt gggaactgtg acaccaaagc ccccaggact atctgcctct 540

ccaggagcca gatagaatga catgcctttt tcctaattgt ccacattcca cccccaagcc 600

actgccactg tgggccaagc catccatctt gcaatcttca tctaaaacag ctctcatttc 660

atgccagttt tgtcaaacct gcaccgtcac aagatattca gaagatgaaa acgttagaag 720

acacccctga attaaaagca cttactagca gggggtggat tatccaaaag tgccgtgatc 780

ggttgtacct gggtttctca ccaatggaat cagctcctgg tgaacaagca tgtgggtg 838

<210> 12
<211> 1033
<212> DNA
<213> Homo sapien

<400> 12
atttaaggct gtacttaact aatttgggct gaggatgaat atatcagcca cagcacatta 60

aagaatgago caaggatttg tcatggttgg tcacttttta aagtatttga ttactgcaac 120

tggagaatga aaagtgtata ttggtgacgc caacctcagt ttctgagcac tcctgctctg 180

tggtgagaat cagacaaaaa ttcatcgggg tgaaaaaaaa aaggcattac ctgattcaca 240

ccctttgtctt gctagccctc ttccattcat ttctcacaca gcactttgct ctgttaaatc 300

ctctctctgt ctcagaccat tgcttgcccc ttcaaagggt atggttcagg ctcctttcaa 360

gacatttgga gtttctctct ggggaaagag agccccctac tggtttggct tcagtctagg 420

tccaccatcc ctctcgatct ggcatcttgg agattaattt aaaaggcaag ctcaccacaa 480

tgtaagccta tggtctggcc aaccttgctt ttgggaactg tgacaccaaa gcccccagga 540

ctatctgcct ctccaggagc cagatagaat gacatgcctt tttcctaatt gtccacattc 600

caccccaac ccactgccac tgtgggccaa gccatccatc ttgcaatctt catctaaaac 660

agctctcatt tcatgccagt tttgctcaaa cctgcaccgt cacaagatat tcagaagatg 720

aaaacgtaga agacacccct gaattaaaaa cacttacata gcagtggctg gaattactcc 780

aaaacgtgcc cagtgatcgc actgtaacat gggattttct cacccaaata ggcaactcat 840

gcttcctgag tgtaatcaaa gcatgtggtg ttttggggcc atatgcacca ggtttctatt 900

ttagaaacct tcagctgtct tgcttatgta ccgtatgtaa atttattctt tttaaaaatc 960

acttttattt gattttgact tattaaatgc tttaaaagcc aaaaaaaaaa aaaaaaaaaa 1020

aaaaattggt cgg 1033

<210> 13
<211> 824
<212> DNA
<213> Homo sapien

<400> 13
acatatgaaa gtgacctcca aggggattgg tgaatagtca taaggatctt caggctgaac 60

agactatgtc tggggaaaga acggattatg ccccattaaa taacaagttg tgttcaagag 120

tcagagcagt gagctcagag gcccttctca ctgagacagc aacatttaaa ccaaaccaga 180

ggaagtattt gtggaactca ctgcctcagt ttgggtaaag gatgagcaga caagtcaact 240

aaagaaaaaa gaaaagcaag gaggagggtt gagcaatcta gagcatggag ttgttaagtg 300

ctctctggat ttgagttgaa gagcatccat ttgagttgaa ggccgcaggg cacaatgagc 360

tctcccttct accaccagaa agtccctggt caggtctcag gtagtgcggt gtggctcagc 420

tgggttttta attagcgcat tctctatcca acatttaatt gtttgaaagc ctccatatag 480

ttagattgtg ctttgtaatt ttgttgttgt tgctctatct tattgtatat gcattgagta 540

ttaacctgaa tgttttgtta cttaaatatt aaaaaacact gttatcctac aaagaaaaaa 600

aaacaaaaca aaaaaaaaca agggtggggg taatccgtcg ggcataagct gtcccctgtg 660

tgactggggt tctccgtccc catccccatc tccgtgacac acaaaaaagg ccagacggcg 720

cacatgcact caccgaccag tacaaccaca caaggaaacg acgtaggcac accaacaacc 780

gacagcagaa agcaatacag aaaacaacaa caaccacacc tcac 824

<210> 14
<211> 1093
<212> DNA
<213> Homo sapien

<400> 14
acctagagtg gttggaccat cagatgtttg ggcaaaactg aaagctcttt gcaaccacac 60

accttccctg agcttacatc actgcccttt tgagcagaaa gtctaaattc cttccaagac 120

agtagaattc catcccagta ccaaagccag ataggccccc taggaaactg aggtaagagc 180

agtctctaaa aactacccac agcagcattg gtgcagggga acttggccat taggttatta 240

tttgagagga aagtcctcac atcaatagta catatgaaag tgacctccaa ggggattggt 300

gaatactcat aaggatcttc aggctgaaca gactatgtct ggggaaagaa cggattatgc 360

cccattaaat aacaagttgt gttcaagagt cagagcagtg agctcagagg cccttctcac 420

tgagacagca acatttaaac caaaccagag gaagtatttg tggaactcac tgcctcagtt 480

tgggtaaagg atgagcagac aagtcaacta aagaaaaaag aaaagcaagg aggagggttg 540

agcaatctag agcatggagt ttgttaagtg ctctctggat ttgagttgaa gagcatccat 600

ttgagttgaa ggccacaggg cacaatgagc tctcccttct accaccagaa agtccctggt 660

caggtctcag gtagtgcggt gtggctcagc tgggttttta attagcgcat tctctatcca 720

acatttaatt gtttgaaagc ctccatatag ttagattgtg ctttgtaatt ttgttgttgt 780

tgctctatct tattgtatat gcattgagta ttaacctgaa tgttttgtta cttaaatatt 840

aaaaacactg ttatcctaca aaaaaaaaaa aaaaaaaaaa aaaaaaacaa gggtgggggt 900

aatccgtcgg gcataagctg tcccctgtgt gactggggtt ctccgtcccc atccccatct 960

ccgtgacaca caaaaaaggc cagacggcgc acatgcactc accgaccagt acaaccacac 1020

aaggaaacga cgtaggcaca ccaacaaccg acagcagaaa gcaatacaga aaacaacaac 1080

aaccacacct cac 1093

<210> 15
<211> 428
<212> DNA
<213> Homo sapien

<400> 15
atgtagcgat ggcatggtca ctaatctgct cacggcgcag tgtggatgga tggtcgcggc 60

gaggtgtttt tgctgtctta atttactaaa ttgcctaatt tttagttgat ttctctacag 120

agcttacctt gtgtagctat ttctctttta ctgttgcttt ggtgtgttct tgaagtttga 180

ttgattattt ttggtcattg agttcttagt gtagtttgct agatgcatta ttaagattct 240

tttgtgtctg acacttgtgg tgagttgctt ttgaaatctg ttgagatttc catgtgaaga 300

actgataatc aggctcttga tctgcttccc taaattactt ttttagagcc caaaccccag 360

gtttagggtg agggtttgta atacaaagaa acacttccat cacttctgcc cttaaccctt 420

gcctctat 428

<210> 16
<211> 6823
<212> DNA
<213> Homo sapien

<400> 16
aaacttggag agacgcagga caggatcccg gcggcagaag gacggagaga aaggggaccc 60

cgggacggga aaggcgcaga gcaggcgcgg gcggcggcgg cggcggggca gggcagggcg 120

ggcgtcccgg cagagggcgc gcggtcgccc tgtcgccctc cgccccgccg gggtcacagt 180

gccccctccc tcgcgcccta gccgccctgc cgggctattt ttacgcgcgg acaccggaca 240

ccggacaccg ggctggggcg gcggcggcgg cggccgaggc ggccgaggcg gggccgcacc 300

ggggccgggc gtcggggcca cacgtcggtt cgcgggtcgc cggggctgcg cgcgccatgg 360

agccgcggtg cccgccgccg tgcggctgct gcgagcggct ggtgctcaac gtggccgggc 420

tgcgcttcga gacgcgggcg cgcacgctgg gccgcttccc ggacactctg ctaggggacc 480

cagcgcgccg cggccgcttc tacgacgacg cgcgccgcga gtatttcttc gaccggcacc 540

ggcccagctt cgacgccgtg ctctactact accagtccgg tgggcggctg cggcggccgg 600

cgcacgtgcc gctcgacgtc ttcctggaag aggtggcctt ctacgggctg ggcgcggcgg 660

ccctggcacg cctgcgcgag gacgagggct gcccggtgcc gcccgagcgc cccctgcccc 720

gccgcgcctt cgcccgccag ctgtggctgc ttttcgagtt tcccgagagc tctcaggccg 780

cgcgcgtgct cgccgtagtc tccgtgctgg tcatcctcgt ctccatcgtc gtcttctgcc 840

tcgagacgct gcctgacttc cgcgacgacc gcgacggcac ggggcttgct gctgcagccg 900

cagccggccc gttccccgct ccgctgaatg gctccagcca aatgcctgga aatccacccc 960

gcctgccctt caatgacccg ttcttcgtgg tggagacgct gtgtatttgt tggttctcct 1020

ttgagctgct ggtacgcctc ctggtctgtc caagcaaggc tatcttcttc aagaacgtga 1080

tgaacctcat cgattttgtg gctatccttc cctactttgt ggcactgggc accgagctgg 1140

cccggcagcg aggggtgggc cagcaggcca tgtcactggc catcctgaga gtcatccgat 1200

tggtgcgtgt cttccgcatc ttcaagctgt cccggcactc aaagggcctg caaatcttgg 1260

gccagacgct tcgggcctcc atgcgtgagc tgggcctcct catcttttc ctcttcatcg 1320

gtgtggtcct cttttccagc gccgtctact ttgccgaagt tgaccgggtg gactcccatt 1380

tcactagcat ccctgagtcc ttctggtggg cggtagtcac catgactaca gttggctatg 1440

gagacatggc acccgtcact gtgggtggca agatagtggg ctctctgtgt gccattgcgg 1500

gcgtgctgac tatttccctg ccagtgcccg tcattgtctc caatttcagc tacttttatc 1560

accgggagac agagggcgaa gaggctggga tgttcagcca tgtggacatg cagccttgtg 1620

gcccactgga gggcaaggcc aatgggggc tggtggacgg ggaggtacct gagctaccac 1680

ctccactctg ggcaccccca gggaaacacc tggtcaccga agtgtgagga acagttgagg 1740

tctgcaggac ctcacacctc cctagaggga gggagggagg gcagggtgga gggcaaggct 1800

gggggggaggg gattgggttt aggaagagct aggttaagtc gtaacgagtg gggaaacact 1860

gagtcttgtt gggtcttggg ttgtgtggtt tggtagctcc tgtgggtacc tcctgaagca 1920

gcagcgaatg gcaatgggtt gtgttgtgtt aatgaagact caattggttc atattactct 1980

gagttgtgca aagctcatgg agccttttgg ggtagtgttg agataggttt ggtcgtatca 2040

ttttgtgagt ttcctaggtc agtgttgggt ttggttgggt tgtgagtctg ggatagtgtg 2100

gtccagctgc attgtgtagg attctgtggt ttggtgggtc ccctagggcc atgttgggtc 2160

aagttagatg gtcccccatg gcattgttga gatcgaatgt gtgtggtgtt aagtttcgtt 2220

gagacatggt ggaaattgtg tagctctgtg attcttccag gggcatgtta ttttaggttc 2280

tgtgaacttg cgagtcatgt agaaatgtga agagtccagt ggtagaattt gagctttcta 2340

ggtcacattg ggttaagttt gtatgaccaa atgaatcttg tagggttctg ttgggcttaa 2400

ctgtgtagag gtgtgtggct ggacattttt cgtggccaca gcgagttgag ttgtgttgaa 2460

ttgtacaacc atatgagcct tgtaaggcca gttcagttgg gtcatgccac tgtttgagtc 2520

tcatagggcc atgctgaatt gagttccatt gagttgtgtc actatgtgag tcctacagga 2580

agttgggttg agttggactg tgcgaacgag ttccataggg ccacatcggg ctgttttgca 2640

tttagtggta gcaccaggac ccaaaggaaa tagcagtggg gaagcatcat gtatctggga 2700

gcatgcagtg gcgagggctc tgggaggtgt gccgagctgg ctccccagct cgctgtaggg 2760

ggcgggactg gattctgtat ccatgggatt gggtgttcat ccagaggcga ctgggtaaat 2820

taggaagagg tggatgctcc tcctgtttac cccacatcca cttcattgtg ctgttcactc 2880

ccattctccc ctacagtttt atgctcagac atggaggtca gagccacaag ggaaagggga 2940

gaggggggaga aaactgtact ctgtccagac atgatagagg gacagagcca aaaggataga 3000

gaaagagacc cagaaaaagg aagaggtgga aacccagaga gacagagacc caaagggaga 3060

gaaacagaga ctcagggaga gggagacaat gacctggagg gtggggtatg gcagagacgc 3120

agaagagagg aacagaaatc cagagtgggg agacagagac caagagcagg ggatagaagc 3180

cgggcgaagt ggcccatgcc tgtaatctca gcactctggg agaccgagga agggggattg 3240

attgaggcca ggagttcaag accagcctgg gcaacatggt gagaccccat ctctacaaaa 3300

aatacaaaaa ttagctgagt gtggtggcac atgcctgtga tcccagctac tcaggaagct 3360

gaggcagaaa gatcccttga ccctgagagg tagaggctgc attgagccat gattgcacca 3420

ctgcactcca gcctgggcaa cagagggagc ccccgtctca acaaacaaac aaaaagagcc 3480

agtgggggag ggagggacag agacccagag ggcagcgtca gacacccaga gttggagaca 3540

gaacaacaga gtctcaggga aagagaacca caatagaaaa aggcagaaaa ggccgggcgc 3600

ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggtgggcaaa ttacgaggtc 3660

aggagatcca gaccatcctg gctaacacgg tgaaaccctg tgcctactaa aaattcaaaa 3720

aattagccgg gtgtactggc atgtgcctgt aatcctagct acttgggagg ctgaggcagg 3780

agaatcactt gaacctggga ggcggagctt gcagtgagcc aagatcacgc cactgcactc 3840

cagcctgggc gacagagcga gactctgtct caaaaataaa tgactagata aaaaaaaaaa 3900

ttgttggagt acattttccc tgcattttag agattaggac agtggtcact ttctagttgc 3960

cgtcttcaca gacaaacttg aaaggtatta aggaaaaaat tccacatagc tctgggcctt 4020

ttttgtattc tgattgccac gttaatctgt aatcgttaag tacccatgaa agaatgtgtt 4080

tcagtatcag tgctttttg gtactttaac aagtattgta ggcacggtgg atttttttga 4140

aatacccatg tgtccaatag caagttaagg aacctagcgc atcaaatttt gttttcttca 4200

ttcatttgga accttatttt cgcaaaaaac ctagctggaa taataatgcc atattgtatt 4260

ttgcacactg ctttattttc tagaggctct gggagcaaat tacattcatc acattacctc 4320

tgctctctca agattaaagt ctttcagcaa cattctttat tgtcatcata aggaaaaact 4380

tttatcgcaa tgatagtcag ttcacattag tttagcacta tagatagtat tcagaaattg 4440

ctttctcttt ggtaaagaag ataaataggt gaattttatt gcctgtgttt tcaggcttca 4500

gaggtacctg gcaccttact caagcaaaaa caaaaaacag ctttacttca aaagcatgtg 4560

cctctggtaa attatccatc acttttgttt taaagtggtc cctccaagaa taatgtggca 4620

tacacagtgg gtttggttgt aaggcagatg taatagatga cttcaaaagc ctgctcctta 4680

ctgcactctg gacaaggccc tggtctgtat agtttggctt cctctgttaa gtcattctac 4740

aaggtgactt gccatttgaa ctttcgtaag agacctcagc agttagactg agaaaagtta 4800

gagtgctgta gttatgaagt catacatttt tgataaaata cctgtggaaa accagggatc 4860

ctctcctgtt tctctctctt gcctcccccc ttccctccat tcctctgtcg cctccccctc 4920

tcctccactt tctttaagaa tcaggggggta cagaataacc tacaaacaac ctcttcaaag 4980

taccccttgga aattgaagtt caatcttaag ttttttgctg tctttaattt actagaattg 5040

cctaattttt agttgatttc tctacagagc ttagcttgtg tagttatttc tcttttactg 5100

ttgctttggt gtgtcttgaa gtttgattga ttatttttgg tcattgagtt cttagtgtag 5160

tttgctagat gattattaag attcttttgt gtctgacact tgtggtgagt tgcttttgaa 5220

atctgttgag atttccatgt gagaactgat aatcaggctc ttgatctgct tccctaaatt 5280

actttttttag agcccaaacc ccaggtttag ggtgagggtt tgtaatacaa agaaacactt 5340

ccatcacttc tgcccttaac ccttgcctct attacctaag cttagtaagt agttttctta 5400

atattgtcca atgtttggtg ttatctgcat tgtttttttgt ttttttaatt gtggttgttc 5460

tgacgtgtga aatttaagga aagagccatt ttttaaaaat gccgagcagg gcaatgcaag 5520

tacagccaaa tattagactt gatgtgcaat cttcggtcct ttttaatctg gggtattata 5580

ggcagtactt taaattgcaa agtcttccgg gcctattttc ctctacattt ttgtaattaa 5640

ctctgggggc ttacttgttt tggcagtact gaaatcaaag gagctggttc ttcttttctc 5700

ccaattattt tcatatgaaa gcacctacaa ttagcctgtt agtcctattc agatacatca 5760

aatatcagtg aatgctttac tattcgcaca tttaagcatc tttgttttac ataaaattag 5820

agtatgaaaa ccagtgttca attttttatc ttgttgagct tgtaaaatgc cagcaattta 5880

aaactaggac ttttcccccc ataagccaag gaggtagaat tactaataca agggttaaag 5940

aaggtagatt ttgttttcaa tatttgggta atattagaaa gattcttccc acagggaaga 6000

actagcaagt gtcccaattt tttccaaacg ttggggaggg gaaaattcac tgtatcatga 6060

aaccctaagg gtttgttgca cttcctgctt tttaggcctg gataacagta tcaccatcct 6120

tatttacaga agggtaaaac tgactcttaa tgagaaaagc tttataagtt caagggctgt 6180

aaaatatgaa ctacttaggg tcgtttgcct tccatgggaa cttggctaga cttagaaaaa 6240

gctgtttgtt gtgctaatgt aaaagtgtca tacaatttag aagatttttg aagatggtaa 6300

acttagaaga attctatgtt ctgaaatgca cacttttaga atgtttttct ttgaaaacag 6360

gctaatagtt ctttcttttt ttgacaaagt ttcagctcct ctttaaagtt attgtgtcat 6420

ttttctggtt taaatttccc ttatatttcc acctgtaatg tcagtggcaa caacatcata 6480

cttcacttac ttgttaattt acttttttacc cttattctaa gagcaagttg agttgaactg 6540

aatcttcctt gtcttagtaa caatgtataa atagtggctt ttctgtacaa aaggttgtaa 6600

tgcctcctga tggatataat tttgtgattg tatttaaaag ttgaataaat cacaccagct 6660

tcctgaaaat gttcataatg catcttttgg aaaacaaata catgcctact ttgtgcatat 6720

ttgcattaac atggcaaaga ttgtatgaaa tacctgtttt tcagaaaata aaggttcagt 6780

ctcaaaagat aaaaaaaaaa aaaaaaaaaa aaaaaaggcg gtc 6823

<210> 17
<211> 984
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (211)..(456)
<223> a, c, g or t

<400> 17
tggccctttg agatttccta tctcaccgtt acttcagttt acccttgcag ggggccaggg 60

agtcaagaat ataccgtgtt cctccagggt ttaagccggc catgccttcc cgagagcata 120

accaacttga caggggtgcc cagttacccc acaaactgaa ggaaggagat ccttcccccg 180

tccccaggag tgctctcaac cagcctcaga nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnncccc aaccggccta ccacgacccc 480

actgaacagc gcatgacgcc acgagaccgg aacgacaagg cggaccaaga cagaaatcag 540

actcaaacac aacagaccaa gataaaccgg caaatccacc acagaaatcc agggcagaaa 600

aaccgacagc aacagcacaa actcgtgaga ccaaaagagc gacacaaaag ggacaaaggg 660

aaacacgaac agaaaagcca gcctatgacc agcacgacac aaccccaaac gccaccagag 720

cacccagacc gagccataga caacaacacc gacagccata acgtacggac tgctgccgac 780

aacccaaccc agcacaaata aagacgcacg gaaccccaca tgcagagcct ccactgaatg 840

gcgccacccg cacggacacg acagcaaccg ccgcaatcag accaagcact ccccaacaga 900

ccacggagag acccgcaaca agcggcagca caagacacag atcggacaac accaagaaga 960

taggagagag ccccaggtaa caac 984

<210> 18
<211> 186
<212> DNA
<213> Homo sapien

<400> 18
ggtcgcggcg aggtacaaac gtgctgtaat tataagtctg tggtgcagtt ccattttatg 60

ctacacatta aggcttttaa ttttgtgtaa aaggaatttt tggaatagtc ttcattgaga 120

ctctttttaa tcttatactt tctaatttag aattataaga cattcaattt gaattataga 180

aaaatt 186

<210> 19
<211> 418
<212> DNA
<213> Homo sapien

<400> 19
gatttgacta tatagccatg gtccttaatc atgctgagcg gcgcagtgtg atggatcgcc 60

cgggcaggtc acattggaat ttgaatgcag tggccaggac agcagcttga taaaccacct 120

tataggtagg taagcaagcc acgggattcc ctcggctggg ctggtgatgg ggaggggcgc 180

agtggacagg aagcgggcag ggcctgagga ccctgtactt taatgggcca caacagggct 240

cagctcctct cagaacttct gctacaatca gagttaagac cagatacatg ctacctaagg 300

ccatgaactt gacaatatct gctccagaga agccctgagc cctcgcccta ggtccctatg 360

attataaacc caatggtaag cccaagtctc acctttctaa atattctgtt tatcagtt 418

<210> 20
<211> 1811
<212> DNA
<213> Homo sapien

<400> 20
atcccgctaa tctgggggcc caatctcccc ccaggggctc ttccaagaca ccgaggcttt 60

cgctcgtccc tgtttaacag cctgcggtgt cttttatata tacgctccct gggggcccgt 120

gtcaccacaa cgtgtggtag tgaaatagtg gcgcctctgg aggaaattcc ttgatggcgc 180

cggcctcttg tatgtgttag aagacgctct ttactatggt gtggtctagt actaattgct 240

gattcatcga tgcggccttt ttttttttt tgacggttcc tatataacgt ttatttctgg 300

aagttaaagt agatacagca atataccaaa aaaaaaaaaa aaaaaaaaaa agacaaaaaa 360

cctcacaata atataaattt ttacactatg aagtacacat tggaatttga atgcagtggc 420

caggacagca gcttataaac caccttatag gtaggtaagc aacccacggg attccctcgg 480

ctgggctggg gatggggagg gggcagtgga caggaagcgg gcagggcctg aggaccctgt 540

tctttaatgg gccacaacag ggctcagctc ctctcagaac ttttgctaca atcagagtta 600

agaccagata catgctacct aaggccatga acttgacaat atctgctcca gagaagccct 660

gaaccctcgc cctaggtccc tatgatttaa acccaatggt aagcccaagt ctcacctttc 720

taaatattct gtttatcagt taagcttatt atgtactgaa gagtgttgct ggagaagtaa 780

gagctctgca tgacacatca actgtctcca ggcaggagga ctggggcgaa ccctgttcat 840

cctagagaag gtcgtctccc ccttaattca gagaccagct aatttgattt aacaaacaga 900

attccacaga aactctgatc agctgaggca aggtggtttt gagttgccaa atccggccat 960

gcctctgagt cagaggcagc ccatccagca cgtgctaggt gttcccatac gcacaggaga 1020

ggcgagctag ccagccaagg cgggcaggcg gggaggccct ctagctgttc tgcctcacct 1080

gtggggcccc agcagggagg agtcaccagc ctcagaggga ggccaggtat acaccctcat 1140

acgagggcag actcaagttt atcagtaagc ctttgccctg catgaactga atgaagacgc 1200

cattacaagt gccactggat atcaccaaga aactggctaa gaaccatatt cctgagctca 1260

accagacacg tcgctggggc cccatagtgt gcatcatgtc caacctgtaa ctctctcccc 1320

ctcttcttcc atgaggtcct gagaccagga ttcctaaaac aaacaggatg agggaccttt 1380

agacacgcaa ggagacatgc ctctagcagg atcaggggga ctggggtcgg gagggtgggg 1440

caggaaggaa gccagaatca gaagtatccc agccctgata aggcacccac agaaacaacc 1500

tagcctctga aacagcagat caagtccagg gagaaaggga gtcaaggttg taagcaccct 1560

ctgcccagcc ttactcacta aaaggccaat acattacaaa ggaaaataag tctatttata 1620

aaaaaaagtc taaaatgctt agattctcct taaaccttcc tatttcaaca ccaaaggcca 1680

ggttgcagct tcagatgtct taagggtttg acctgaagcc cgtttcaggg attatataaa 1740

ttaccagcct gaaggaagct gaatgcttaa agcaccaagg aggctgttaa ctgaagtttc 1800

ttttttttt t 1811

<210> 21
<211> 602
<212> DNA
<213> Homo sapien

<400> 21
cggccgccgg caggtacagg ttttttggt aatgcgattt aaaattttta aaaacatggg 60

cttcatgaaa acaccattgt tactaacttg ggcttagatg ggaatctaat gtgaatagtt 120

caagttatgg gacttgtcta aatgtgtcta tactctaaac ctgggggaag gtggtgtgga 180

acattgactt gactacagtc gaatacatgt tgggaatcat tcccaatcat aagacatcgt 240

cgtttctgta ggatgactgc atatgttcag agtagctttt tgaatttggt tcccttagct 300

ttagggatgt gatgttatag tccaaaatgt ttacaagaaa acctaagtct tcaaaagcac 360

aacttttgtt gcttaggact ttgcatcaac tgttgttcca gacctcactt caacttcttg 420

gtctctgaac tggttttagc tagcatgcat gagagacagt tttcatgtat gctagctaaa 480

accagttcag agacagtttt catgtataat gtttctgctc tgcaactgga aagtggttta 540

ctttgcacaa atgctgatga aaacttcatg actatgagac ttttcttctg tgtatcaaga 600

aa 602

<210> 22
<211> 4114
<212> DNA
<213> Homo sapien

<400> 22
ttttttttttt aattaggata atgcctttat taacgaaaat gaaacgttca ttcctccttc 60

cactcctttt cgttggtttt ttggacacag ctcacctgat cctgctaaaa acgttgtcag 120

tctgcttggg gcttccctcc ttgattgact cacgctgggg gatgtcttga aaagtattta 180

tccacttcat gggaatgagc cctccaatat cagccaacat caatcattct tacctaaaga 240

ataataagaa aaagttaata taaaagacaa gggtataaaa taaaggtttg aaaatgctag 300

tcaacttcaa aatttaaaga gtaaaaatcc agagataaag attgggggta agttacagca 360

taaaaaaata ggaagaaact tcatggtggg ggggaaatct aaaattattc ttacataaaa 420

taagtagaca cctgaattag aatgaaaact gtattttctt taaaatgtaa aagcctgact 480

ctcagtttca ccagtctgag cacaagtttg actgcaaccc aaaatatact atcccttatg 540

tgaaggtatg tgacaacgtt gacctcacca aatgagtttt aacatcagct ctttttttcat 600

atgaaagcac ataccctgct ccccattcaa gtatgtcttc cattgtcagg caggctgacc 660

accttcagca ggagtcctcc aagagtgccc aactcccctt cccacagtac acaacgctgt 720

agttgttgtc ctgcaatcct ttgtatttac ctcattcttt cccatctaag tcctcactga 780

gttttaaagt tagggctgga aaagctatgc cttactggga cagcaaggaa ccaatttttt 840

tctgagggag aagacattca ccttcactat atgcctggca gggccacagt gcacaaaaca 900

aagatcagcc ttcattcaag ttccaggttt ttcttcctcc ctgaatgatt actgcaaagg 960

gtatatgaag taagagttcc ctgttgcaca tgtaccatcc ataagggata ctatatcgtt 1020

ttgcattctt ccccccattc tccacattgt cctatcttaa gtccaagccc ttttcactct 1080

caaaaaaaaa aaaaaaatat ttttttcagc actggtgttc aaaagcaacg tttttatggt 1140

taatggttta ccagcaactg ttgagatttc cagttgagtc ttaaaaattg ccaatcatta 1200

tctagcagca atgacagatg attaggagca gtcaaatcct ctgaattctt tccctaatag 1260

gcagccattt gagaactgca ctagctgaca tcactaaaac attatcagct aaagccaaaa 1320

ccaaataaag gcccagacca acatcctggc tctctaaaac ctgtccaaaa tcattaagtg 1380

aaaggcagta aatgcaggac tgtggatcat gtcactgcag ctgacaatga ttaacaatag 1440

gagacatgca acccccatta aggttaaaag tccaaaacta gtcacacgca tctctttatt 1500

ggggaaaagt gagactatta tgcattcttg gtaggtttgc aaccttgcat gaagagcacc 1560

cattgcattt ctttcatctt tcagaaagca ccggtatctg ttccaagggc ctaacagtac 1620

gaaaatacat tctggcatca cacctctgaa cccaagactg ttctcattaa aaataatttt 1680

ggtttgtaac aaaattatga aatacaatgc aagcacctcg gtatagcatt attactgaaa 1740

ccacttaatt cccagctttt tgagtttttt aaaaaaaccc actgcactaa gattcacaat 1800

tcattgctac atacaaatta aagctagtaa gaacacacta acgtcacaag tttctcattc 1860

taaagtgcaa aagcctaatc atctgaaagt gaacagggta aggcaaaatt aacccccccac 1920

cccaataaag ttcctgaagt ccatatatta tataccaagt acattctcta aaaattgtta 1980

ctgactggta agaaatagac ctgagttttt atttctaaca cccaatcact aaaccacggc 2040

agcaagcact ggccaccgat ttaatggatt acgacacagg aaaccccatc agggttctat 2100

gtaatttagt gatactcatg tcactaatat tgagcattat acttgatctg cattatattg 2160

ttgatatgca gaggctaaac tagtcatcat ttgctctttc atctatcagt agagtccaaa 2220

gttgtttgct tgaatggact acatgttaaa gtacaagtct gtccccacct tgtgaattgc 2280

ttgccaacga gcaagctttt tcttgataca cagaagaaaa gtctcatagt catgaagttt 2340

tcatcagcat ttatgcaaag taaaccactt tccagttaca gagcagaaac attatacatg 2400

aaaactgtct ctcatgcatg ctagctaaaa ccagttcaga gaccaagaag ttgaagtgag 2460

gtctggaaca acagttgatg caaagtccta agcaacaaaa gttgtgcttt tgaagactta 2520

ggttttcttg taaacatttt ggactataac atcacatccc taaagctaag ggaaccaaat 2580

tcaaaaagct actctgaaca tatgcagtca tcctacagaa acgacgatgt cttatgattg 2640

ggaatgattc ccaacatgta ttcgactgta gtcaagtcaa tgttccacac caccttcccc 2700

caggtttaga gtatagacac atttagacaa gtcccataac ttgaactatt cacattagat 2760

tcccatctaa gcccaagtta gtaacaatgg tgttttcatg aagcccatgt ttttaaaaat 2820

tttaaatcgc attacaaaaa aacctgtact tttagttcaa ctcaacttgt agaattacca 2880

agattgcata atgaaattac tgatattgcc gatctatggg caggtcagtt tgctacaata 2940

gagactaatt atcacatgct atacggtcca tgtcaaggtg ctaaaagcac cctagttccc 3000

aagtatagtt tagttccctc tcccccacac cactgatgtg ttccatgtta tcttcagtta 3060

caatgcaact aaaggaaacc acaactgagt cagatatacc aaagaatcaa gttgcacttt 3120

ttatctgaga actgcaacag cactgaattc tgcctgacaa attacagctc taaccccaca 3180

cccacacagt tttgatgtaa gctagcttta ccatacaagt gttaggtgct gcactgtaat 3240

ttcatgtcag aaatgtgatg ccagaatgcc cacggaataa aagtacatac aagtcaccaa 3300

gttagattat atgcttgtta cctacctgta tgcagtcggc aatgagaatc ttggagcaag 3360

caggaatact acatccgggt cctaatgtcc attgccattt gcggtactac gttcctcaca 3420

gttacgcact gcagaaatgc tggctaaatg cagttatgta gcaggccact actttaatag 3480

tgcataattg cagtccaaga acaccagaaa acattccgcc acaacttagt ggcttgccca 3540

agaaaagcca agtatctaaa ttttaatctg ccataatatg ccacttaaaa attgcacagg 3600

cgtaacatta caatttcccc atttttttagc tgtttatatt agtggtacaa tacatctata 3660

aagagtggtg gggttaggtc tgtaatttgt caggcagaat tcagtgctgt tgcagttctc 3720

agataaaaag tgcaaacttg attctttggt atatctgact cagttgtgtg gtgtccttta 3780

gttgcattgt aactgaagat aacatggaac acatcagtgg tgtggggtta ggtctgtaat 3840

ttgtcaggca gagtgaggtt tgttgtggag ctggcagatc caaagttgga ggtgaaatgg 3900

tataaaaatg gtcaagaaat tcgacccagt accaaataca tctttgaaca caaaggatgc 3960

cagagaatcc tgtttatcaa taactgtcag atgacagatg attcagagta ttatgtgaca 4020

gccggtgatg cgaaatgttc cactgagctc ttcgtaagag agcctccatt tatggtgccg 4080

agcagctgga tagaaacccc cgctgattgt tgtt 4114

<210> 23
<211> 234
<212> DNA
<213> Homo sapien

<400> 23
agtgtttgca acagcaccat ttgtcaaatt caaagatgct caaaaggtgt tccctacttt 60

gcatgagagg gagagctttg taacaggaaa ttgtataagg caaactctct attcattcct 120

aaggcctctg ttcattccta atgtttacat ggttctctac tctgaagggc accaacatgg 180

acctcacctt cttaacatgg aaaatcaaaa tctaaatgaa tacaattaaa agga 234

<210> 24
<211> 600
<212> DNA
<213> Homo sapien

<400> 24
actgcattgg tggcttctcc atagggaggg ggtcctatct catagactcc aaacttcatg 60

atggcaggga ggatatctgg atttgctcac ccctgtattt tcagagcctc acacagtgcc 120

cggaacatgg taaacattta atactatctg tcaaaaacaa atgactgaaa gagcagatgg 180

aaagagccaa tcctgcattg aagagatatc catggtggcc ctaaaactac tcaaaccaga 240

tgtctcttca gcttcccact ggaaaatgga ccggtgggca aaccaccatc tcacatctca 300

gcgagagggt caatgtgcaa aagtatttaa atagcgacga ggtcccacac cccatcccag 360

gacagccccc agcatcctgc tccaacataa gcagaaaaag acaatgaggt gcaaagagct 420

tatctcctgg gatggttccc gggtggtctg cattgccaca cagcacatca tgctttgatc 480

ctgaaggaca aggccccaat aagccctatt tctcagcaca gtcttaaaat cacctggaaa 540

tcattccacc taggacccct cgtgaatcag cactgatccc aagcgacacg ggaaccccta 600

<210> 25
<211> 496
<212> DNA
<213> Homo sapien

<400> 25
ggcaggtaca ttgtaataat aaccacaaaa taattgtata aggagaaaaa taataactga 60

cattcatggc cctctgcaag gcacagtatc tgctttctca tttggtcctt cgttggttcc 120

atcctttgaa ggtctgagag cagcaggccc attccatgaa caccaaagcc cttcctacca 180

cgccagccca gactgccatt tcccctccag aaggccagtg ttcatcatcg atagggctag 240

agaccatccc ggagtcccca tgtttcagga ctcccgaatc ttcaaatagt ccatctttaa 300

gaagggatct cctggcagct aagagggtga agctaatcgt tctacaaagt agtgcctaac 360

tgttgataag atggcagtgt gtaggaagct gtgtgttggg tcggattccc cttttcatga 420

gccattttct gtggtaggtt ccacggacat ggacaccatg gcctcatgga agcatgaaca 480

gctccaacac agggtg 496

<210> 26
<211> 1690
<212> DNA
<213> Homo sapien

<400> 26
atgcctagca cggatgctgg tgcctcgcag ggtttacgaa tgatctgtcc cgtctccagt 60

gacgccctct caggacccac taccgaagtc attaggactc ctttgtttgc agatgacagt 120

taccgagctc agataagagg tggcaaagag tgccagtccc ccatcaaggg tggacccaac 180

aggtggacag ctcagagagg cctggatgcc ggggatgggc tctcagcaga tgtaggagga 240

gctcaagtgc tggcaacagg caagacccct ggggctgaaa ttgatttcaa gtacgccctc 300

atcgggactg ctgtgggtgt cgccatatct gctggcttcc tggccctgaa gatctgcatg 360

atcaggaggc acttatttga cgacgactct tccgacctga aaagcacacc tgggggcctc 420

agtgacacca tcccgctaaa gaagagagcc ccaaggcgaa accacaattt ctccaaaaga 480

gatgcacagg tgattgagct gtaggtgagc agtgacgtga agaggggttc tagccccgtg 540

gaaaacagcc catggttaac atctcaggat gttctgcatt caaacaccca aggctggtaa 600

tgaactttca catggactga atattggagg caaataatag aaggaataga atatacagtg 660

cctctgtcct gaaggaaaat atcatgcctc ttctggaaga aacggactgc acagaggaag 720

gattgagcaa tttagcctgc agtggaagaa ggtggacacc aaaagcttca ccctgtgttg 780

gagctgttca tgcttccatg aggccatggt gtccatgtcc gtggaaccta ccacagaaaa 840

tggctcatga aaaggggaat ccgaccccaac acacagcttc ctacacactg ccatcttatc 900

aacagttagg cactactttg tagaacgatt agcttcaccc tcttagctgc caggagatcc 960

cttcttaaag atggactatg tgaagattcg ggagtcctga aacatgggga ctccgggatg 1020

gtctctagcc ctatcgatga tgaacactgg ccttctggag gggaaatggc agtctgggct 1080

ggcgtggtag gaagggcttt ggtgttcatg gaatgggcct gctgctctca gaccttcaaa 1140

ggatggaacc aacgaaggac caaatgagaa agcagatgct gtgccttgca gagggccatg 1200

aatgtcagtt attatttttc tccttataca attattttgt ggttattatt acaatgtaca 1260

tggctgttgc atagaagaca tgactggtgg aggctgagga aagccatgac attctacaat 1320

tgccatcagg ctaaggcccc gtgagcattt ctctcccttg taatattaac cctgtatttc 1380

tgggatcaca tcacggaata ttctttgcct ttccactttc caggaaatct ctcggactgg 1440

gctaccctcc ttgtgtgtga tgaaagatga gctatatttc agaacaaagt gctgtgttgt 1500

catgatttgc ctggactccc agggcgtctc ttacccaact tgataacgat gctgttcatt 1560

agcagccttt gttaactgat aaccaagagc ggtaatgtga tactcataag caattttctg 1620

tgtgtaggat aaaataaacc atcttgtatg ggaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1680

aaaggcggtc 1690

<210> 27

<211> 461
<212> DNA
<213> Homo sapien

<400> 27
gggtgataat catataggcg aatggtctct agatgctgct cgagcggcgc agtgtgatgg 60

atgcgcccgg gcaggtacca agctcccagg cccttcctct atcatggatg ctgggtgact 120

tcgggaagtc accacctctt cccaagcctg tttcccatat cacagatgtg gggccatggc 180

ctcgatgatg gtctccacag gtctttccac ctctgtgagt ccaagtcagg tcaatcagca 240

aggacccatc tctgccctgg gtcagctcct cagaaccaac ccccagcatc cctaaagcaa 300

aagcctcacc tcaagggctg ctcagaagag agcaccttca gcatgagttg ttgctggaag 360

atctaataag ctgtgtttcc tgggaagtgg tgctttactt agccctgtgg acaacttctc 420

tatgcatctg tgtgagcaga tgatcattgt attacctttt a 461

<210> 28
<211> 4043
<212> DNA
<213> Homo sapien

<400> 28
ccgcttaatt aaagatcttt ttttttttt ttttttctc ataaacagga actttattaa 60

actacatgtt acataaaaga acatataaat ggaccattaa atacattcag tttattttaa 120

acaaatttac atagatactt atttacattt ctccattgta ttcttaaatt atttttccaa 180

gcttactacc gataaaaggt aatacaatga tcatctgctc acacagatgc atagagaagt 240

tgtccacagg gctaagtaaa gcaccacttc ccaggaaaca cagcttatta gatcttccag 300

caacaactca tgctgaaggt gctctcttct gagcagccct tgaggtgagg cttttgcttt 360

agagatgctg ggggttggtt ctgaggagct gacccagggc agagatgggt ccttgctgat 420

tgacctgact tggactcaca gaggtggaaa gacctgtgga gaccatcatc gaggccatgg 480

ccccacatct gtgatatggg aaacaggctt gggaagaggt ggtgacttcc caaagtcacc 540

cagcatccat gatagaggaa gggcctggga gcttggtacc cagggttcct ccaagaggtc 600

ccgatccctc tgctatccac aaatccacaa acttagaatc acttgcatcc attttccacc 660

accatggacc ttatgctttg aactgttttg tctacctgat taaatatata acttcttagg 720

cttcttgaga tggtgaaaag cagctcagga ttcccataag caactttgtg gagccctgga 780

aatacctact cagggctgtt tttacaagag gttttgtggt caggtgcttt actacttcag 840

ccataacgtt tacctttaaa actcagctga cttatggaag ctcagcattt ccaattcgct 900

tagatgacag gcaacagtct gcagaaggga ggttctaacg tcaaccacgt ggattcccca 960

caaacgcatc atatttgcct aaatccatct atctaccaat gtcagatcta aaatgaggtt 1020

tcactaataa gtgacctaaa attaaaaaaa acacaaaaaa gtttcttaaa aaaaaaattc 1080

caagaatttc cccgtttccc aattagtctc agaggagtcg tgaaatgggc attgggtaga 1140

aaaagtccat gttttcaact gtttcttttg aataaagcct gattccccca ctctgccccc 1200

aaacttttcc cgatgatata cattcttctg aaagcaattt catgctttta gtctgcttta 1260

aaaaacctga cttggttctt ctcaggattt aaaaaccact gtaccaagga agagatgtcc 1320

acgttaaaaa ctattattaa gagacagaga caaggagaga gagactctct ccaggccatg 1380

ttagaaaaga aggaagtgct gggagcacac aaagagagaa gaaaggctgg gagacagaat 1440

ggtttgcacc tccctgtaac cagagtgaac acaggtcata gcagccttca ctcagtttta 1500

taacagacac acgaatggag gaacatggga ttggaaagac gcacactgtc aaacagccct 1560

gatgcctttc cactaaaata tggcacttct agattccaga gaaatccagc attctaccag 1620

gggctgccac ccatctgcag ggaagcctgc cctgcaacca tgaggccctc aaggctcata 1680

ctccactgct cacaatgcgc gccctgaata catggaagca tctgatttct gtctgagtcc 1740

aactttccct cctgttccta gacagactca cacccagagt catgcacgga gaagtgggca 1800

agccagacaa tcttcaggga gacacagccc atcacaggat acccaaggct tatggggaag 1860

tccagttgcc agcaacggac cctgtgtggg caggtgaaac tctgaagcca gagaacacag 1920

gaagataaaa atatcttcat actgaggata tactgcacaa gtgtggtggc tcacacctgt 1980

aatcccagca ccttgggagg ccaaagtggg ctgatagctt gagcctagga gtttgaaact 2040

agcctgggca acataatgag accctaactc tacaaaaaaa aaaaaaatac caaaaaaaaa 2100

aaaaaaatca gctgtgttgg tagtatgtgc ctgtagtccc agctatccag gaggctgaga 2160

tgggagatca cctgagccca caacctggag tcttgatcat gctactgaac tgtagcctgg 2220

gcaacagagg atagtgagat cctgtctcaa aaaaaaaaat taattaaaaa gccaggaaac 2280

aagacttagc tctaacatct aacatagctg acaaaggagt aatttgatgt ggaattcaac 2340

ctgatattta aaagttataa aatatctata attcacaatt tggggtaaga taaagcactt 2400

gcagtttcca aagattttac aagtttacct ctcatattta tttccttatt gtgtctattt 2460

tagagcacca aatatatact aaatggaatg gacaggggat tcagatatta ttttcaaagt 2520

gacattattt gctgttggtt aatatatgct ctttttgttt ctgtcaacca aaggatggac 2580

agtgattcag aaccgtcaag acggtagtgt tgactttggc aggaaatggg atccatataa 2640

acagggattt ggaaatgttg caaccaacac agatgggaag aattactgtg gcctaccagg 2700

taacgaacag gcatgcaaaa taaaatcatt ctatttgaaa tgggattttt tttaattaaa 2760

aaacattcat tgttggaagc ctgttttagg cagttaagag gagtttcctg acaaaaatgt 2820

ggaagctaaa gataagggaa gaaaggcagt ttttagtttc ccaaaatttt atttttggtg 2880

agagatttta ttttgttttt cttttaggtg aatattggct tggaaatgat aaaattagcc 2940

agcttaccag gatgggaccc acagaacttt tgatagaaat ggaggactgg aaaggagaca 3000

aagtaaaggc tcactatgga ggattcactg tacagaatga agccaacaaa taccagatct 3060

cagtgaacaa atacagagga acagccggta atgccctcat ggatggagca tctcagctga 3120

tgggagaaaa caggaccatg accattcaca acggcatgtt cttcagcacg tatgacagag 3180

acaatgacgg ctggtatgtg tggcactctt tgctcctgct ttaaaaatca cactaatatc 3240

attactcaga atcattaaca atatttttaa tagctaccac ttcctgggca cttactgtca 3300

gccactgtcc taagctcttt atgcatcact cgaaagcatt tcaactataa ggtagacatt 3360

cttattctca ttttacagat gagatttaga gagattacgt gatttgtcca atgtcacaca 3420

actacccaga gataaaacta gaatttgagc acagttactt tctgaataat gagcatttag 3480

ataaatacct atatctctat attctaaagt gtgtgtgaaa actttcattt tcatttccag 3540

ggttctctga tactaagggt tgtaaaagct attattccag tataaagtaa caaacacagt 3600

ccctagatgg attgccacaa aggcccagtt atctctcttt cttgctatag ggcacaggag 3660

gtctttggtg tattagtgtg actctatgta tagcacccaa aggaaagact actgtgcaca 3720

cgagtgtagc agtcttttat gggtaatctg caaaacgtaa cttgaccacc gtagttctgt 3780

ttctaataac gccaaacaca ttttctttca ggttaacatc agatcccaga aaacagtgtt 3840

ctaaagaaga cggtggtgga tggtggtata atagatgtca tgcagccaat ccaaacggca 3900

gatactactg gggtggacag tacacctggg acatggcaaa gcatggcaca gatgatggtg 3960

tagtatggat gaattggaag gggtcatggt actcaatgag gaagatgagt atgaagatca 4020

ggcccttctt cccacagcaa tag 4043

<210> 29
<211> 176
<212> DNA
<213> Homo sapien

<400> 29
ttataataac aaaaaaagtg tgaggggatg ttttcccaag ccccttctc cggtgggcgg 60

ctctcgtgtg aggacccatg gtgacagagt ctctctcatc tcctcactct gaaagcattc 120

cactggggag agtcaaccct ggctcggggc ttcctccaca cagcacacgg cccttc 176

<210> 30
<211> 1332
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (623)..(642)
<223>

<220>
<221> misc_feature
<222> (623)..(642)
<223> a, c, g or t

<400> 30
atcccgggaa cttgacacct cgtcccgggg cagcagcggc tctcctgcgc acgccgagtc 60

ctactccagc ggcggcggcg gccagcagaa attccgggta gatatgcctg gctcaggcag 120

tgcattcatc cccaccatca acgccatcac gaccagccag gacctgcagt ggatggtgca 180

gcccacagtg atcacctcca tgtccaaccc ataccctcgc tcgcatccct acagcccccc 240

gccgggcctg gcctctgtcg ctggacacat ggccctccca agacctggcg tgatcaagac 300

cattggcacc accgtgggcc gcaggaggag agatgagcag ctgtctcctg aagaggagga 360

gaagcgtcgc atccggcggg agaggaacaa gctggctgca gccaagtgcc ggaaccgacg 420

ccgggagctg acagagaagc tgcaggcggt gaggaactct gcgtagggtg ggagcacctc 480

tgggtgggct ggagtgagag ccccgggggt cctgatcctc tctcccacag ctgtctcctt 540

acccacaaat gccctacaga tgagtcagtg gagatagagc tttccttcct ttggacacat 600

gggtcagcat tgtgttgatt gannnnnnnn nnnnnnnnnn nngtggcgtt tagcaccagc 660

cagggcccag agggaagaaa tgaacaacct tctgtggggg tgctagccac ggtttttacg 720

cagcacctga gtgttctctg ctcacagaac tgtgggggct cctgggtgag ggatgatgga 780

gggtgggagc cagggcaggt ccagacctgc cccatccagg gctgcttccc gaggttccag 840

ccactcccat gcctgccgcc atcaggagtg gcccagcctg gcctaagcac tctgggttgg 900

gccagagaga agtcttgatg gagataggtc acgtggcagc agggttgtat tttcttgggg 960

ttttctcatt tcctttatgg gcaggcgggt tccccacctc tatcatgaac tcagagctgt 1020

ggccatggta ccgcgtgctg ctcttgtgag tcactctccc gtgccgccta attgcccaga 1080

cctggggacc cgcagccctg cccttccctc tgttaagctt gacctgacca acgtcacgta 1140

actcccttca gagcctggcc ctgcaggaag ggccgtgtgc tgtgtggagg aagccccgag 1200

ccagggttga ctctccccag tggaatgctt tcagagtgag gagatgagag agactctgtc 1260

agtggtcctc acagtgactg cccaggatgg ggcttgggaa aacatcccct cacacttttt 1320

ttgttattat aa 1332

<210> 31
<211> 571
<212> DNA
<213> Homo sapien

<400> 31
gccgcccggg caggtacctg tagtcccagc tactgaggag gctgaagcag gaggatccct 60

cgagcacaaa agtttcaggc tgtggtaagc tatgactgtg ccactgcaca ccagcctgag 120

ttacagaggg agatcccaac tcttaaaaac taaaacaaca ataaatatat acaagaatca 180

taacataaag ggattcatgc ttagaaaaaa tccataaact cccttctaaa tattgagaca 240

ctccaggctt ctttcagaca aataacttct aattattcca tatttttcaa gttattaacc 300

aagataaaga atctctcagt tagtggggaa aatgaaaatt attaagaata gaattgtctt 360

ctgactttaa aaacaattta gactttaaaa catgaactgt ttactcaggc tggtgatact 420

ctagtttgtt agtataccat acttgaagat atcatcaaga tcactatagt tgtatatatt 480

ctctattttt atatgtaatg ttaacttagt tcaagtattt tttgcttgta tcgttaactg 540

atcatcaaat acaatcctaa agatatatca g 571

<210> 32
<211> 240
<212> DNA
<213> Homo sapien

<400> 32
ccgggcaggt acatattcta cagaaaaaaa tcaaatatgt atacaaataa gtatgcacag 60

gatttagaat cttacataaa aatgtattta acttggctgg agtgtgtgtg tgtgtttccc 120

aggttaagta aaattagaaa gccagaatca caagccacaa aaaagaaaaa ctgataaatt 180

agataccatc aaaattaaaa tctctaactg ctacaaataa taccatcaag aatgggaatt 240

<210> 33
<211> 1026
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (883)..(883)
<223> a, c, g or t

<400> 33
aatagctcga ccgcgcgttt tgatggatga gcgacgaggt caggaggacc tgtatcacaa 60

aaaaaaagaa aagggaaaaa agaagagaaa gcagcagcat gatacctgac atgacagatg 120

tgggagaccc acagcctgca gacactgtgt ggctggaagg tggcgacggg agtggtgccg 180

gtggagtgtg gagctgtctg aaagtgacgc agcagacagt agaagcatag gtgggcgaag 240

cccaggtgac cctcaaaacg ttgcacaaga acatcaggga aaaagaacta gaatccttta 300

aggaaaatgt tcttcatgta tgagagacta aagtgatttt tctaagaaag ttcagccctt 360

ctctgactta cctggacatt tctagatact tccaaaggac cctctgggaa tccatagctt 420

cctaatctgg agatgggagg tcataaggga gacgctgtgg ggttccttga agtttcttgg 480

gttcacagag gagccccctc acttggtgtt ctcccgtgag ccagcctcca cctgccaaag 540

acactctggt cctcgtatag tgagtaatgg ggctcagggc ctctccaaca acagagagga 600

gctgatgctg tagggctgac cccgtgactt cctggagtcc tcaccctgtc cagtgctttg 660

agattcttcc cacctcccca tcctcaccag ccggatcggg cgctgtgcag tgtggtcagc 720

atggtgaaga aagtcatttc cttggtggac agtattcctc tttatctctc attacactgg 780

aaatgttatt tctgctgtat catccgtgct ccaacgtttc tagtctgtca ggctcacctt 840

ctctctggaa agaattgctt aacttgacat tccatgtgcc gcntataaaa tatattttga 900

aagaacaaaa aaaaccaaaa aaaaaaaaac ggtgggggta acctgggcca aaacgcgtcc 960

ccggggtgaa ttgttctccg ccccactcaa ccccaccaac aaaaaaataa acagacaaaa 1020

aacaaa 1026

<210> 34
<211> 1545
<212> DNA
<213> Homo sapien

<400> 34
gccgacaatg acatcggagc cgtctcaacc acagggcatg gggaaagcat cctgaaggtg 60

aacctggcta gactcaccct gttccacata gaacaaggaa agacggtaga agaggctgcg 120

gacctatcgt tgggttatat gaagtcaagg gttaaaggtt taggtggcct catcgtggtt 180

agcaaaacag gagactgggt ggcaaagtgg acctccacct ccatgccctg ggcagccgcc 240

aaggacggca agctgcactt cggaattgat cctgacgata ctactatcac cgaccttccc 300

taagccgctg gaagattgta ttccagatgc tagcttagag gtcaagtaca gtctcctcat 360

gagacatagc ctaatcaatt agatctagaa ttggaaaaat tgtcccgtct gtcacttgtt 420

ttgttgcctt aataagcatc tgaatgtttg gttgtggggc gggttctgaa gcgatgagag 480

aaatgcccgt attaggagga ttacttgagc cctggaggtc aaagctgagg tgagccatga 540

ttactccact gcactccagc ctgggcaaca gagccaggcc ctgtatcaaa aaaaaaagaa 600

aagggaaaaa agaaagaaag cagcagcatg atcctgacat gacagatgtg ggagacccac 660

agcctgcaga cactgtgggc tggaaggtgg gaagggaggg gccggtggag gtggagctgt 720

ttgaaagtga cacagcagca gtagaagcag tggtgggcga agcccaggtg accctcagaa 780

cgttgcacaa gaacatcagg gaaaagaacc agaatccttt aaggaaaatg ttcttcatgt 840

atgagagact aaagtgattt ttctaagaaa gttcagccct tctctgactt acctggacat 900

ttctagatac ttccaaagga ccctctggga atccatagct tcctaatctg gagatgggag 960

gtcataaggg agacgctgtg gggttccttg aagtttcttg ggttcacaga ggagcccccc 1020

cacttggtgt tctcccgtga gccagcctcc acctgccaaa gacactctgg tcctcgtata 1080

gtgagtaatg gggctcaggg cctctccaac aacagagagg agctgatgct gtagggctga 1140

ccccgtgact tcctgagtcc tcaccctgtc cagtgctttg agattcttcc cacctcccca 1200

tcctcaccag ccggatcggg cgctgtgcag tgtggtcagc atggtgaaga aagtcatttc 1260

ctcggtgggc agtattcctc tttatctctc attacactgg aaatgttatt tctgctgtat 1320

catccgtgct caacgtttta gtctgtcagg ctcaccttct ctctggaaag aatttgctta 1380

acttgacatt ccatgtgccg ctaataaaat atattttgaa agaaaaaaaa aaaaaaaaaa 1440

aaaaaaaacg gtgggggtaa cctgggccaa aacgcgtccc cggggtgaat tgttctccgc 1500

cccactcaac cccaccaaca aaaaaataaa cagacaaaaa acaaa 1545

<210> 35
<211> 338
<212> DNA
<213> Homo sapien

<400> 35
tgatcactca ctataggcct ggtgctctag atcatgctcg agcggcgcag tgtgatggat 60

ggccgcccgg gcaggtactc agaatggagg ggctggaagt ggggccagtg gtcctggtag 120

tcagcctccc cttctgacac aagaattatc gtggaacagc ttgtgtgacc gatcaactgg 180

tctctccatc tttaagccat tgtcttgttg actctgttac tgcagagttt ggggaggtct 240

catggcttct cgggatccta tttcctgagc ttccctgaag ccaacttctg gccactggta 300

tgtgatgcga gcacagtgcc tcctcctgct cctccagg 338

<210> 36

<211> 1851
<212> DNA
<213> Homo sapien

<400> 36
gccgcatatt ttttttttt ttttttttt tttttcaagg ttggaaaatc ccactttact 60

tcaagtgaag tcactgggtt gtccccatga aatatctcca caatgggtgt gctagacctt 120

tactctctgt ggcaccacca agacccctc accctgccag agccctctga agtgtactca 180

gaatggaggg gctggaagtg gggccagtgg tcctggtagt cagcctcccc ttctgacaca 240

agaattatcg tggaacagct tgtgtgaccg atcaactggt ctctccatct ttaagccatt 300

gtcttgttga ctctgttact gcagagtttg gggaggtctc atggcttctc gcatcctatt 360

tcctgagctt ccctgaagcc aacttctggc cactggtatg tgatgcgagc acagtgcctc 420

ctcctgctcc tccaggaagc ttctgctgat tgaacgcagg ccagatggcg agatccggag 480

caaccttctt agatcaaggg aagaaagggg cacaagaggg gagtaggtaa caagataaaa 540

ggagctccct ccctgatgac tacagagcca tcgtggcagc cctgggcctc catttcagac 600

gttcatcttg cctaagccac caccatcagg gtctcagtca gtcatcattc tcatttactc 660

gggtggctgg gtgagggcgg aatactacct tccagctgtc tgagattaag cctaagccac 720

caccatcagg gtctcagtca gtcatcattc tcatttactc gggtggctgg gcgagggcgg 780

aatactacct tccagctgtc tgagattaag cagaacagca gctaaagcag taacagcagg 840

tctccttcag cagcttgcca cagggaagag ggtcccgcgt gaaccgaact caacttccac 900

ctgcgcagag gtagctacca tttgcgtaga ggtagctggg tctttaaggt tagagggaag 960

gcgcagggag ggagagacgg cgggtggggg ttgaacaaag tggagattca caaaagcaga 1020

ctagggcggg cgacgtgatc agatgacctg tgcgggcggc agcctcctgc cctcctcccc 1080

ttcgtgcgcc ggctggagcg aagagttctt ttgacagccg tgagcttccc cgccaggaac 1140

ttactggggc tgcatcaccc tagaaacgtg gctttgggct gtggaaacgc tgcctctgtg 1200

gaagtctctc ctcgcggggg tggacgggtc gctgcgcgcc cagcgttctt ctgcggttct 1260

cacagcccgc cgccgccgcc gcctcgggga cctttgcggg gaggcctcag gtcagcgccg 1320

cccttgcgat ggcgggagag cagagcggcc ccagggcctc tgagctccgg ccccggcacg 1380

tcccgccctc tcctccgctt gcgggagccg gggcgcctcg gagggaaacc ttccggacac 1440

aggccgggag aagaggggcc gtggcgccat cgcgcagctt cctggtttcg gccatgtttt 1500

ttttttttg agacggagtt tcgctctttt tgcccaggct ggagggcaat ggcgccgtct 1560

cggctccctg cagccgccgc ctcccgggtt gaagggactc tccggcctca ggccgcggtg 1620

ttcttgtaaa ttagaaccgg cctatgcgcc aagcgcgggt ctcgcggctg cggagagagg 1680

cccagcaggg ggcggggacc ccgggaggcg gggcctggcc gagctgcgca cacccgggtg 1740

gttctggcta cgggtacccg gacctgccca acgaccgttc tccattcccg acctccgcac 1800

cctacccggt ccgcagggca ccttcctcct ctctctcagg cacgtcctcg a 1851

<210> 37
<211> 409
<212> DNA
<213> Homo sapien

<400> 37
gagagcaagt tcaagcaggg ccaggaacag gacagccggc aggagagcag gctcaacgag 60

gactttctgg gaatgctggt ccacaccagg tccctgctga aggagacact ggacatctct 120

gtgggactca ggaacaaata cgagctgctg gccctcacca ttaggagcca tgggacccga 180

ctaggtcggc tgaaaaatga ttatcttaaa gtataggtgg aaggatacaa atgctagaaa 240

gagggaatca aatcagcccc gttttggagg gtgggggaca gaagatgggg ctacatttcc 300

cccataccta ctatttttt atatcccgat ttgcactttg agaatacatc taaggtcatc 360

tttcaaaaga gaaaaattgg acacttgagt gacttgtttt tagtttgtt 409

<210> 38
<211> 2112
<212> DNA
<213> Homo sapien

<400> 38
ttgagagcag agagatcctc aggatatctt tagccaaagg aaaagctccg cattcccacc 60

cagtccagaa attgaaatac tatcaggggg caagagcctt tctctccagc tacacactcc 120

atctcccggg agcaagggga aactccgaga ggagggcaac agagccagca tcttgccagg 180

gccccggagg aggggttccc cgctacgcct gtgccggagg agttccagtc accgagcgag 240

gggcgcaagg gtgggtgcat cctgcgctgc ggcgggcgcg ctacccagac gctggtgtgc 300

agagccacat gaagcctgct ggggactggg ggccagggag cagcaagcca gctgggactg 360

aggcggacgc tgtctcaggg agacgctgac tcgcaaagac actcccttcc ttgtgcctgg 420

gtaaaaagtc tcctcctggg gtccctggcc atcctgaata tccagaatgg tgtttctgaa 480

gttcttctgc atgagtttct tctgccacct gtgtcaaggc tacttcgatg gccccctcta 540

cccagagatg tccaatggga ctctgcacca ctacttcgtg cccgatgggg actatgagga 600

gaacgatgac cccgagaagt gccagctgct cttcagggtg agtgaccaca ggcgctgctc 660

ccaggggggag gggagccagg ttggcagcct gctgagcctc accctgcggg aggagttcac 720

cgtgctgggc cgccaggtgg aggatgctgg gcgcgtgctg gagggcatca gcaaaagcat 780

ctcctacgac ctagacgggg aagagagcta tggcaagtac ctgcggcggg agtcccacca 840

gatcggggat gcctactcca actcggacaa atccctcact gagctggaga gcaagttcaa 900

gcagggccag gaacaggaca gccggcagga gagcaggctc aacgaggact ttctgggaat 960

gctggtccac accaggtccc tgctgaagga gacactggac atctctgtgg ggctcaggga 1020

caaatacgag ctgctggccc tcaccattag gagccatggg acccgactag gtcggctgaa 1080

aaatgattat cttaaagtat aggtggaagg atacaaatgc tagaaagagg gaatcaaatc 1140

agccccgttt tggagggtgg gggacagaag atggggctac atttccccca tacctactat 1200

ttttttatat cccgatttgc actttgagaa tacatctaag gtcatctttc aaaagagaaa 1260

aattggacac ttgagtgact ttgtttttag ttttgttttt gtacattatt tatgtgattg 1320

ttatggaatt gtcacctgga aagaacaatt ttaagcaatg tcatttctag atgggtttct 1380

aattctgcag agacacccgt ttcagccaca tctaaaagag cacagtttat gtggtgcgga 1440

attaaacttc cccatcctgc agattatgtg gaaataccca aagataatag tgcatagctc 1500

ctttcagcct ctagccttca ctcctgggct ccaaaagcta tcccagttgc ctgtttttca 1560

aatgaggttc aaggtgctgc tttgcatgcc tgccaaccca tggaagttgt ttcttacttc 1620

ttttctctct tatttattaa ccatggtctg agagttgttt ttgttctatg taacagtatt 1680

gccacaaaac tataggcaaa tcgtgtttgc agggagattt ctgatgcctc tgtgggtgtg 1740

tgtaagttaa agtggccaca tttaagaagg ccaagctttg tagtggttgc acagtcacac 1800

tgatatgctg atttgctctt tctcattgta tgtctatgct ttgtcatcag tgctatagta 1860

aattacaaag aaataggtag attgtatgaa catacccaca aatgcctatg atttaggtta 1920

ccaatgtatt ctttctcatt tggggttttg cttctgtctg tctgtttatt ggaaacttgt 1980

acttcaagta gggggaatcc taattctaat aactccttag ctaagtttta ttattcaggc 2040

aataaacatg ttttcatgta aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2100

aaaaagatcg gc 2112

<210> 39
<211> 713
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (260)..(539)
<223> a, c, g or t

<400> 39
ccgcccgggc aggtacctac ctgatgaagc tgttcatgct tccagcatca aactggacgt 60

ttactgtctg cccaaaagcc agtttccaaa aggtttgctt ccctctgttc agtggattct 120

tgactacata taggtcatat atttcaaaaa ataatgccta gctatttcta ctttgaaatc 180

atgactaaag ccaaaccaca accacagcaa agataaccta aggatttgtt taccagaaat 240

acctacaaaa aagtttgcan nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnc 540

caccctgcca aaaagtttgc catggtccgg caggcaaagt tgcaaccgga gccgttcctg 600

ggtgaaaagc ttggcggtaa tccttgtcct atccttgttc ccttgtgtgg acatttgttt 660

tccccgctcc ccaatctccc ccataccctc ccacaaaaaa aaaccaggac aac 713

<210> 40
<211> 338
<212> DNA
<213> Homo sapien

<400> 40
ctgatttaaa gtcatagaat ttactgaaat acctacacag ctctgaatcc aacatttttg 60

tcatggttat tagaagaaac caaaacaact gactgaaaat gaaatgttca gtcagacagt 120

gtttgtccac atctttttt ttttgaagtt ttgaatcatc cattcacaga aaacacttgt 180

ttatgggtat aatgcaaaac ttttgaaaca aaaaaaccta ctaaaaatgc tttcgctaaa 240

gtgattggct tttcattcat gctttgaaat aaaattatct agaaaggttg gagaagggtt 300

tgccgaacag ccatcttcct gatgtgcctt agattaga 338

<210> 41
<211> 805
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (241)..(520)
<223> a, c, g or t

<400> 41

aatgctgctc gagcgcgcgc agttatgatg gatgtggtcg cggccgaggt ctcagtttcc 60

ctatttgtaa aatggaaata atggtagctg cctcaaacct cattacgaat tcaatgagtt 120

aaacttgaaa agacttatta cagtagctgg cacataaaga cttgatagta gtttatatgg 180

atgctatctc atattagcat atatggaatt aatagtgtat cactatactt ttttttcttt 240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ttactatgct atcaattcca 540

tatatggaac ggtgtcttcc ctttcagagc tctttaaagc tctgcagaag atttacatgt 600

gtttatagag ctaagagaaa tcgggccatg gaaatgtggg tgtgtaacta acaaatttaa 660

cctcttcatg ttatattaat catgcataat tctatcttca tcctctgaaa gtggcctaga 720

gcaaccatca cagtaggaag cttgggtatc atggccatag ctggtcctgg tgtgaattgt 780

ttcgtcaatc caaaaaaaaa aaaag 805

<210> 42
<211> 300
<212> DNA
<213> Homo sapien

<400> 42
gaatgtatga tcactatagg gccatgggtt atctagatgc tgctcgagcg cgcgcagtgt 60

gatggatcgt ggtcgcggcg aggtaccccct tgagcctggg caaggtgcat gacacacact 120

tggtcactgg aggaaatgat gagtagagag tgacaagcag aggaggggag ctttgggcct 180

gagtcctctg gggcacaaga gtggttgagg cttggcactt gccacctaga tttcagacaa 240

tgtgtcagga atcctggatg cccagaaaga aacctgctca gggtctggag ccccaacata 300

<210> 43
<211> 561
<212> DNA
<213> Homo sapien

<400> 43
ggactgagaa ttcctgtggt ctcctgagat cagtttcttt ccaccagatt gtaaacaagc 60

tgtggaaagt tttagttctt agcgactcca cgacccacac tagtcaggca gaagccccca 120

tgtacaaggt acccccttgag ccttggcaag gtgcatgaca cacacttggt cactggagga 180

aatgatgagt agagagtgac aagcagagga ggggagcttt gggcctgagt cctctggggc 240

acaagagtgg ttgaggcttg gcacttgcca cctagatttc agacaatgtg tcaggaatcc 300

tggatgccca gaaagaaacc tgctcagggt ctggagcccc aacataaagc ttttatttag 360

gcaatgttga gcactaatat gtggttggag ctctgcaggg ggtctccatg ggtgcactgt 420

gtagcaaagc tgtgggagaa aggccttacc cctcaaaatt ccagaattat aaatctatca 480

gccgcttgca acttcagcct ggaaaagcca tgggcattaa actccaaact gtgagagcag 540

ccatgtggct gcaccctcaa g 561

<210> 44
<211> 530
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (102)..(182)
<223> a, c, g or t

<400> 44
gttttaaaaa gaggagctcc cctgcatgag atctcacttt ttgcctgcta ccatttatgt 60

aagatgtgac ttgctccttc ttgccttcca tcatgactgt gnnnnnnnnn nnnnnnnnnn 120

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180

nngcccgccg cgggcgtctc cacccgcctc tccgccgccg ccccaccccc gccgccccgc 240

gcgccacgag cagatcggag cagcaggccg cgaggagcag gcgagcggcc gccagtaggc 300

gcgacgcgtg ccgcggagac gacgcccggg gcgtgcacgc cgctgaggcc gaggacagcg 360

atgcgccgcg cgaggcgcgg agcgaggagg agcggcgtgg accgccccag cggccgagga 420

cgggcgagcg gcgcggggag gcccgaaggc gaggaggagg tgcgggacgg ggggacgaga 480

agaggagaag agaggggaaa gggagggggg ggagcggaga agagggaggc 530

<210> 45
<211> 709
<212> DNA
<213> Homo sapien

<400> 45
acaaaatagt aatgatagcc atggaagttt taccttattc tgtgagaagt gttcttaaac 60

ttattaagtg tctaaactaa ggtttagtgc ttttttaaag gaaagttgtc ccaggattca 120

tcctaaagaa agcaaaagtt aattcaactg atccaccaat ggaattagat gggtagagtt 180

gggttcttga gttttaccac cacttagttc ccgctgaatt ttgtaacttc ctgtgtttgc 240

atcctctgtt cctattctgc ccttgctctg tgtcatctca gtcatttgac ttagaaagtg 300

cccttcaaaa ggaccctgtt cactgctgca cttttcaatg aattaaaatt tatttctgtt 360

ctagctggga acacaaacaa caactcaact acaaaatcac gctctgtggg cagtatccct 420

tgcgccctta gcctttcccc cgggtgtcac tttgtcttcg cccccccatt cccccccact 480

ctcccacact ctccctcctc ccatcacttc cctccctctc tcccgctctc gctgccggcc 540

cgtcagcgcg cgtcgcgctc gcgccgtgcg gctcgcctcg ccgtctctcc ctcctctccg 600

ccgcgtcccc gccgccgcgc gtctgcccgt ccctcgtcct ccgctcccct cctccgcccg 660

ccgcctcccc ccgcccgtct gccactcgct ccctagtgcg cgcgcgccc 709

<210> 46
<211> 1808
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (153)..(153)
<223> a, c, g or t

<400> 46
tcataagagt gaaatagtca gctgctgacg gcacctcagc cacgccacty ttactcagtt 60

cagtgggtgt gcttgcgtgg taggatgtgg tgcagccctc tctacgctct tctatttttg 120

gtatatttyc tatttaacct tcaaatagct tcnaattctt tttttcttgg actggcttca 180

ttctgaattt gtgctaaaat aatctttcat aaagagacct cagtttatag cgtaacagac 240

tacacaatgc actgatgttt tcataatgtt taagggaccc actgcaagaa ggttgctgcc 300

tccttttaat tgtattcatt tagattttga ttttccatgt taagaaggtg aggtccatgt 360

tggtgccctt cagagtagag aaccatgtaa acattaggaa tgaacagagg ccttaggaat 420

gaatagagag tttgccttat acaatttcct gttacaaagc tctccctctc atgcaaagta 480

gggaacacct tttgagcatc tttgaatttg acaaatggtg ctgttgcaaa cacttttttt 540

ttgagatgaa gtctcgcggt tgtcacccgg gctggagtgc agtggcgtga tctcggctca 600

ctgcaacttc cacctcctgg gttccagcag ttctcctgcc tcagcctccc aagtagctga 660

gattacaggc gcctgccacc ccacctggct gatttttgta attttagtag agacggggtt 720

tcaccatgtt ggccaggctg attaactcct gacctcaggt gatccacctt tctcggcctc 780

ccaaagtgct gggattacgg gtgtgagcca ccgtgcccgg cctgcaaaca cattttaatt 840

gacaacacta gggctgttgt acaaaatagt aatgatagcc atggaagttt taccttattc 900

tgtgagaagt gttcttaaac ttattaagtg tctaaactaa ggtttagtgc ttttttaaag 960

gaaagttgtc ccaggattca tcctaaagaa agcaaaagtt aattcaactg atccaccaat 1020

ggaattagat gggtagagtt gggttcttga gttttaccac cacttagttc ccactgaatt 1080

ttgtaacttc ctgtgtttgc atcctctgtt cctattctgc ccttgctctg tgtcatctca 1140

gtcatttgac ttagaaagtg cccttcaaaa ggaccctgtt cactgctgca cttttcaatg 1200

aattaaaatt tatttctgtt ctagtgggaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1260

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1320

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa actgagaaga gggggtgccg 1380

tgcttgtcct tctccccacg ggcgggcgca gcgcgtgcac cacccccgtg gtgttaagtt 1440

aaaaggggggc agcgggcccg ctcaggcggg tgtatataag agagagagag aggggcggct 1500

tacacacaaa atagatctgg cggcagaagg aggcgaataa aaaagttgct ttcagaagtc 1560

ctcatgatac ttgggctgaa gaagacaaga acatcctttt cagaatgcgg atgtagagat 1620

aatatcaatc gaaggagcca caccacccta agtagaaaat tagtcagggt cctcagtcca 1680

tattgagcag taaaaatatt caatgattca gaagtcgtta cggatctaca acagcaagca 1740

tgatgttact tcctgccgcg gaagactgag actttgctgc tcgagtatct ataagacaaa 1800

taatatcg 1808

<210> 47
<211> 760
<212> DNA
<213> Homo sapien

<400> 47
aatgctgctc gagccgcgcg ttatgtgatg gatgccgccg ggcaggtact gggtcccgag 60

cgtggtgggt atttgccaca ctgggtggcc agctcagcag ccccccacct ctctttattc 120

tctccaaagc tggtctttct gactatcatt gtggtagggg gaggacagat gctaaaggtg 180

gaagctgacc tggagaaaga gacacacggg gtgactgtgg caaaggacag ctggaaaaga 240

aactctatca cttcttcatt ggcaaccaca aggcacccga ggccatggca ctcccagagg 300

ctgtgcgcag tagccaagcc tctcaacctc ttctggccct gcgtcctgca gcgaagtctc 360

tgctgtaaga cagtagactc cttcgatgag gtgctcaaaa atgctacccg gggtggtggt 420

gtctggcttg cagtctggcc cagttcagag aaagttgcag agatcagggg ccaaggatgt 480

catagcccca ggttgtcctc agggtcccaa tcctagggca gggtgtgcat ggaagcaaga 540

actatggaaa cctagctcca gtctgcaggc tctgagcccc tagttcctca ctccagcggg 600

gctccctcac tgcacagaac ccacccttc tgtgtgggca ctgctgacca cacagatgac 660

ccagacccaa agagcctggc agaagctctg tgcgttggag ctgggctccg tcctccggtc 720

tggttcaggg ggatcaggaa ggctcttttc acctgtggct 760

<210> 48
<211> 4940
<212> DNA
<213> Homo sapien

<400> 48
tgaagttcag tagagccggc cgctgcaggc cagaaggtgg gagccagcgg gggcatcgcc 60

gcccgcgccc ctctaagtgc cgggccgcaa gctccaccgc agccgcctgc aagcagcggc 120

gcctcggccc tcgacctgcg cgcaaagcct gtgctggagc cgtcctcccg cggcggggac 180

cgggaccggg gacccaagcc aatcgaaagc tccaaccatg gccatggggc tcttccgcgt 240

gtgtctggtg gtggtgacgg ccatcatcaa ccacccgctg ctgttcccgc gggagaacgc 300

cacagtcccc gagaacgagg aggagatcat ccgcaagatg caggcgcacc aggagaagct 360

gcagctggag cagttgcgcc tggaggagga ggtggctcgg ctggcggccg aaaaggaggc 420

actggagcag gtggcggagg agggcaggca gcagaacgag acacgcgtgg cctgggacct 480

ctggagcacc ctctgcatga tcctcttcct gatgatcgag gtgtggcggc aggaccacca 540

ggaggggccc tcacctgagt gcctgggcgg tgaggaggat gagctgcctg ggctgggggg 600

cgccccttg cagggcctca ccctgcccaa caaggccacg cttggccact tttatgagcg 660

ctgcatccgg ggggccacgg ccgatgcagc ccgtacccgg gagttcctgg aaggcttcgt 720

ggatgacttg ctggaagccc tgaggagcct ctgcaaccgg gacaccgaca tggaggtgga 780

ggacttcatt ggcgtggaca gcatgtacga gaactggcag gtggacaggc cactgctgtg 840

ccacctttc gtgcccttca cacccccga gccctaccgc ttccacccag agctctggtg 900

ctccggccgc tcagtgcccc tggatcgcca gggctacggc cagattaagg tggtccgcgc 960

cgatggggac acattgagct gcatctgcgg caagaccaag ctcggggaag acatgctgtg 1020

tctcctgcac ggcaggaaca gcatggcgcc tccctgcggc gacatggaga acctgctgtg 1080

tgccacagat tccctgtacc tggacacgat gcaggtcatg aagtggttcc agacggccct 1140

caccagagcc tggaagggca tcgcccacaa gtacgagttc gacctggcct ttggccagct 1200

ggacagcccg gggtccctga agatcaagtt ccgttcaggg aagttcatgc ccttcaacct 1260

gattcctgtg atccagtgtg atgactcgga cctgtacttt gtctcccacc ttcccaggga 1320

gccctctgag ggcaccccag cctccagcac agactggctc ctgtcctttg ctgtctatga 1380

gcgacacttc ctcaggacga cactaaaggc actgcccgag ggcgcctgcc acctcagctg 1440

cctgcagata gcatccttcc tgctctccaa gcagagccgc ctgaccggtc ccagcgggct 1500

cagcagctac cacctgaaga cggccctact gcacctccta ctcctccggc aggccgccga 1560

ctggaaggcg gggcagctgg acgctcgtct gcacgagttg ctgtgcttcc tggagaagag 1620

cttgctccag aagaagctcc accacttctt catcggcaac cgcaaggtgc ctgaggccat 1680

gggactccct gaggccgtgc tcagggccga gcccctcaac ctcttccggc ccttcgtcct 1740

gcagcgaagc ctttaccgta agacactgga ctccttctat gagatgctca agaatgcccc 1800

agcgctcatt agcgagtatt ccctacatgt cccctcagac cagcctaccc caaaaagctg 1860

acgtctttta cagaatgtgg gatcctcgag ctaagatgag ggcatccctc acgttcacac 1920

ccctggtggc atctgccagc cctgttctgg ggacaaggcg ggctttcgtg ggagccatgc 1980

tcagcctgcc aggaagccaa gccctacagt gcagaggaaa cagaatttca acgggaagct 2040

ggtttgcttc ataccattgg gatctgctgg taaagctgtt atttgggttt agggactgat 2100

cccttgcagt ttacttctgg atcaccatga atggccaaga tggtggcaga acacgctgtg 2160

gaccctgagt tagagacaat gcaaatgttg gattgggtgt aattcttttt gaatcccaga 2220

tccagtctgt acttgaatat gagcagagga tctacaagaa tgctgacagg gaaccgtgtt 2280

aagacccagc acccctattc ccaggagctt ctggcctgac catctgcagc caaagcacta 2340

acagggacag atatgggaat gtccaccttt gatccgcatc ctgcacaata gtggtcccac 2400

catggctgcc actttttat actatttgga gaaaagacct tgtataaatt cgaggcccga 2460

gtgactaacg tctctgtcac acggaaatgg gtacttggtg gcatagagaa acacaattag 2520

ccactttttc agctacactt ctcactcagc tgcaccctac acttctcact caggtgcacc 2580

cccttctgct gtcctttccc caacgtactg ggtcccgagc gtggtgggta tttgccacac 2640

tgggtgccag ctcagcagcc ccccacctct ctttattctc tccaaagctg gtctttctga 2700

ctatcattgt ggtaggggga ggacagatgc taaaggtgga agctgacctg gagaaagaga 2760

cacacggggt gactgtggca aaggacagct ggaaaagaaa ctctatcact tcttcattgg 2820

caaccacaag gcacctgagg ccatggcact cccagaggct gtgcgcagag ccaagcctct 2880

caacctcttc tggccctgcg tcctgcagcg aagtctctgc tgtaagacag tagactcctt 2940

cgatgaggtg ctcaaaaatg ctacccgggg tggtggtgtc tggcttgcag tctggcccag 3000

ttcagagaaa gttgcagaga tcaggggcca aggatgtcat agccccaggt tgtcctcagg 3060

gtcccaatcc tagggcaggg tgtgcatgga agcaagaact atggaaacct agctccagtc 3120

tgcaggctct gagcccctag ttcctcactc cagcggggct ccctcactgc acagaaccca 3180

ccccttctgt gtgggcactg ctgaccacac agatgaccca gacccaaaga gcctggcaga 3240

agctctgtgg ttggagctgg gctccgtctc caggtctggt tcagggggat caggaaggct 3300

cttttccacc tgtggcttca ctggcccttt gagatttcct atctcaccgt tacttcagtt 3360

acccttgcag ggggccaggg agtcaagaat ataccgtgtt cctccagggt ttaagccggc 3420

catgccttcc cgagagcata accaacttga caggggtgcc cagttacccc acaaactgaa 3480

ggaaggagat ccttccccca tccccaggag tgctctcaac cagcctcaga aagcttgaga 3540

agatggaccc tttgcccacc agggttaatt cctggtgggg cagctcggct gtgatcaggg 3600

caaccaaacc tataggaagc cttccagtgt gagctggaat tagactgaac atgtgcttgg 3660

gcctgcctct ccctagacgc agttgcgggg cactccaggg aatgaaccag ctcaagtgtg 3720

tccctaacag cagcctggag ctacccccaa tccctcacag cctgaccctc ctcattccat 3780

cagatgcatt tgtagaatcg gggcaaattt ccttatttat ttatggccca tgcctttccc 3840

cctcttcatc ctgatcccgt tttgctttga agagacccca gtaaccaaaa aacagcctcc 3900

agaagccaaa accatgcctg gatctcccat agcttctcct ttgcttccag gagaaagttc 3960

actgaaaaaa aaatatcttc tggcttcttg tgtgtacaga gacaacagaa ctcggtgggg 4020

aaacgggaat cttttctgca ccaaagctgc ttctaaagca gaaagcagtg gggctcttgg 4080

tgtttcatgc tgccttattt atattaaagg aagaattaaa tcttgcaagg agtaaaaatg 4140

gtcactgttt ggtttttaaa cttcaggaaa tctgtgaggc tcccagggca gagttgggga 4200

caggggggtg gatttcctat tgacaaagca gaagctttca ctccctttt attcttacac 4260

tttcaggaga ctttaaaaaa taacaaaaca gaacatttct accctgtctt tgaggcaact 4320

tgtgtgctgc ccgcagggct ccagaaaggg cctttgaaag ctggctgtag ggtggattcc 4380

agcctaagac cttatttcac agagctaatg tcagcgagaa atataggagc tggaactaac 4440

aggaagcttt tgggttttaa agaaagagat tgtattttaa atagaggatt ttgagtcctc 4500

tagggcactt tggcctttcc actaaggaca accctgccct ctttccccag agtctctttc 4560

ccaaggatgc ttgactttcc ccagcaactg ctcacttgct ttgcaaaata acaagggtga 4620

aatgaataga gttggccccc aggccaagag gctggttatg ttggggaagg gccttatatc 4680

ctcaaaagct ctagcccctc tgcatgcctg gctgaagcac agccccgcgt tgagaacctg 4740

ggggtggagc tctggttcat catcaagggt gtcatccgaa tcacccagac aacctggtcc 4800

attgccccga ccccgccttg aacctacatg gtgctcagaa tccagtctgg agtggtcagg 4860

ctaggctcga tttttaaaga tttcccccaa gtgtctgatg tacccctagc tgaggaagag 4920

ctcagcttta tggtaccgac 4940

<210> 49
<211> 782
<212> DNA
<213> Homo sapien

<400> 49
accattctga acgatgttaa agcaagtgtg gtttatttat ggcatgaacc atgtaacttg 60

aaatatgaac ttacaaggag gggcactcat taggtaacaa gttcttacac caaacttcct 120

tgatgaaata agccaaaata atcctaaaat tcattaaaaa gaacttgata aaagactcaa 180

ataaatgtta gaaagagccc ataattttag gactcctata aaattcttcc tgtttgttaa 240

tgctattaaa actcagattc aagggaaata ccagcttcca cttgagtcac tttgaaatag 300

ttaattcaac agcaccatgt tagaaatata ttggcagcca agactctgaa ctctgcagaa 360

acatttgttt cacccagact tcaaactcta gccctgacta tgatgcccct gtgtgcattt 420

acaataaaga ctccaacgga gggagcctgt tggtgttaaa ttgtttacat ttctttatac 480

aaataatgtg ctttcagtgc cttagttaca gctccctttc tgtttcttgt tccaaaggga 540

ttctgtagta tgtaggtgtg ttttcttagg taaagtctct tattgctact gaaagggaaa 600

tggtctctaa acactggtca ctgtagcagg taaacactac tctaacgtgg gagaaatgag 660

cctccatgct gaggtagggg gtgccctaaa gcctgttatt tatgctgtga aaacgaaatg 720

ggtttgctac ctgataagct gggggatcca tggcctagct gttcctgggt gaatgtttcg 780

ca 782

<210> 50
<211> 2675
<212> DNA
<213> Homo sapien

<400> 50
ggtttatcct aaattactca acccccttgta gccttgacaa attttacctt aaaaccaaaa 60

tgaaacacaa aaattaatcc ttaataatga tagcaagtga tctttctttt tagttttagc 120

cttccttttt caatagtaat atttaaaccc actcttgacc aattgtttgc ccaaatattc 180

ttgtcatttg gagtcagtgg aaaatccagc acaccaagca ccagtcttct tctgaggcaa 240

aagaaaagtg ttgtcatttt cactctgttg gagctgcaca ctttttttct ttttcttttt 300

tttttttttgt cgtgtaagaa ggatgctggt cagagctgca gaaaatatga ggcaattaaa 360

agtctttagc tgttagcaaa cctgttagtt ttacttctgc attgaaccag cctcagaagc 420

tacttactgc tttatgtact ctttgggcat taatgccttc tctgtaatta tatctcgttt 480

ttgcttgaca gtgacctagc cagtaattgc atcgtgcatt gccatgaaag gtaaacacat 540

tgtgaactga acttaccaag cagattctgt aagaaagcgc tggttgaggc tgaacactgt 600

tgacacatca tttttattgg aagagtatta actggtgcct cttctgaaac acaccaaccc 660

atattcctct gctcccccaa agctgtttct gatcctgctg ggagcaacta actagttatt 720

atgcacatct gctccagacc cagctcttta acttcacggt tttacagctt gtttttttctt 780

tttcttttct tttctttttt ttttttaaaa aagcaccttt ttttgttgtt tcccccttaat 840

aaaagaggtt tctaatttat gtttctctaa gatttccttt ggttgtattt agaaacacaa 900

atagttttat aatcaggatt gcatttgtgc tgggtaaaag aggagagctg tgctttctac 960

cccaaagttg ttggtaatgg agccggtctg cccttcctac aaactcaaga ggctcattgt 1020

tcaaggtgta acaacattta attgcatgtc ctctaacgct ttgaaacctt tagagggaag 1080

atcctcattt tttacaattt tgttcctttt catcatagaa aacatgttta agataaaata 1140

caaactttac ctaccccaaa ttcataaagt acatttagtg ctgtatagca ctaaaactta 1200

gagatacaga cactgtactt actttttaag aattagagac agtaactcca aaaataattg 1260

tcctttttc tttctttttt ttttttttaca ataaggctct tgaaaattgt cattacttgt 1320

gttttcctat acattcatct gtgtgaaagc ctttttcttc tttgatttaa aaaaattata 1380

cagttaatgg tttagaactt agaactactt agaattaatg ctaaagtgtc aggaagaaat 1440

taatttagct tcaataattg tgactggcct caggaattct cccttccaca cctgcccacc 1500

tcacctcacc tcacctcacc gcaccgcacc gcaccgcacc gcaccgcacc agagccagag 1560

cagctgcttg tctgcagcag gacacggttc ctacatacgt ttcagttctt tcatggtaag 1620

ctcaatggac tttgaattgt ttacagtgct gtatgtccaa ttgttaaatg taccattctg 1680

aacgatgtta aagcaagtgt ggtttattta tggcatgaac catgtaactt gaaatatgaa 1740

cttacaagga ggggcactca ttaggtaaca agttcttaca ccaaacttcc ttgatgaaat 1800

aagccaaaat aatcctaaaa ttcattagaa gaacttgata aaagactcaa ataaatgtta 1860

gaaagagccc gtaattttag gactcctata aaattcttcc ttgtttgtta atgctattaa 1920

aactcagatt caagggaaat accagcttcc acttgagtca ctttgaaata gttaattcaa 1980

cagcaccatg ttagaaatat attggcagcc aagactctga actctgcaga aacatttgtt 2040

tcacccagac ttcaaactct agccctgact atgatgcccc tgtgtgcatt tacaataaag 2100

actccaacgg agggagcctg ttggtgttaa attgtttaca tttctttata caaataatgt 2160

gctttcagtg ccttagttac agctcccttt ctgtttcttg ttccaaggat tctgtagtat 2220

gtagtgtgtt tcttaggtaa agtctctttt tgctactgaa agggaaatgg tctctaaaca 2280

ctggtcactg tagcaggtaa acactactct aacgtggaga aatgagcttc atgctgaggt 2340

agtggttgcc ttagagctgt tatttatgct gtagaaaacg aaaatggttt tgctacctga 2400

taagcttcag attagatata gccctaaagt tatcctgtac ctgcattaaa ttttcatttt 2460

agaagagaat cttggttttg gtaattcatt ttttttaag atcattatgt gcaaaacacc 2520

aagttttaaa aataactcac agaatggcct tagttttcaa tgtctgatgg tatatttgtg 2580

agttgtgtta tctgtaatat acatcccaga ataaaggagt gaattaaata gtttaaaaaa 2640

aaaaaaaaaa aaaaaagatc tttaattaag cggtc 2675

<210> 51
<211> 168
<212> DNA
<213> Homo sapien

<400> 51
acaacacaac ctgttaggta gcaggctggg tattacctga gtaaaaagtc cctattcttt 60

ccatttctcc atgctattta aaagcatgtg gtgatccttc aataaattaa agatataaca 120

aattatttga tgccattagg cggcctggat caccaattct aagtcttt 168

<210> 52
<211> 1139
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (291)..(576)
<223> a, c, g or t

<400> 52
caaaattgga ttcttcaaaa cgccaggttg gcctttcttg actccaaaag aggtgactct 60

tgtagcattc ttacaggcct aacatagcac ttactttatg ttgcaatgct gtgtgcaatg 120

ctgtgttttt ttactcacta gctctgagca aaacattatt tattcttggt aattgtaaca 180

cttggtagtg aagacatatt ttttggccat ataatggatt atactttgta tttgaatcca 240

tgtgaatcag tagcatattg gtcttcattg tttttcctcc ttctccgacg nnnnnnnnnn 300

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnntaac catatttttt ccaattatta 600

aaaaatcttt tttcccccta aggaatgtgt ttctaaggtc accatgtttt tagtgcttca 660

gtttgacttg tctacatcta aagttaactc tttaaggctt acatttatga ttataattct 720

ttccatcgag ggatttagtc tccattattg tcatgtttct tgatttttt cattatactt 780

agttcttctt ttgtctaggt tttatatata ccttaattaa ttctttcttg gctgttttat 840

atctatgttg ctgtcagcat gggaatctct attattagaa agctattaac ttttaaaata 900

tgtatctttg gtgatttctt tgtcctaaca gtttagatgg cctgttgctc ttctaaatct 960

taaatcatag gtattctatt ttggttgttg tgtttcatta cactgtcata gtttttgata 1020

ttttcattac tattgaggta ctgtggttta ccaatgttct gttggtacga atactttcca 1080

tctgttgggg agctcaaact gggaaccagc acgttcgtga tgttcgcacc ccccacccc 1139

<210> 53
<211> 681
<212> DNA
<213> Homo sapien

<400> 53
ggtacgggga gaacatatca aaaaggggaa ggatggattc ccttgatgcc caggattaca 60

gtggcaccta aagcacattt ttttttttct gagccaacca gctaaaggat cactgcagct 120

aaatacagat agagaagcaa caaagccagg caaataccca tcagagacag tgacaagagc 180

agctgggggc acgggggagg cggaaggaag agaaagaagg ggaggagcct ccagagtccc 240

agccccaacc ccctctgcca tgggctaccc ttgctcccca aaaatcccgt tggggttgaa 300

gtgaggagga ctgcaggctg gggtgaaaat acacaaggac agcccaacaa aatacaacaa 360

ggactagcat cagtctcccc cttactccga cccccaagaa aaataccctt attggtgact 420

agtatttatg aaaatctgta aggagactat tctatgtagt gggctctaat cccatataca 480

cagcagctgc ctgtgtgggg aacctttca aatcagtgat tgcgggaaca aacagtattt 540

tcagcttctt acggtgccca tgcaggcttt taccaagacc ttggttaagt cccagtcaca 600

tttactttct gtcttcatct agaaaagggt gaggaaagag ggagggacaa ggtgctccag 660

gtgctagtaa gctgggtatc c 681

<210> 54
<211> 3191
<212> DNA
<213> Homo sapien

<400> 54

gtgctgccga gtagtcccgg aagcgaagca gcgatggcgg agagtccgac tgaggaggcg 60

gcaacggcgg gcgccggggc ggcgggcccc ggggcgagca gcgttgctgg tgttgttggc 120

gttagcggca gcggcggcgg gttcgggccg cctttcctgc cggatgtgtg ggcggcggcg 180

gcggcagcgg gctcccgccc tcagccgctg cccacggggc cgcgctgctt agccactggg 240

accccacgct cagctccgac tgggacggcg agcgcaccgc gccgcagtgt ctactccgga 300

tcaagcggga tatcatgtcc atttataagg agcctcctcc aggaatgttc gttgtacctg 360

atactgttga catgactaag attcatgcat tgatcacagg cccatttgac actccttatg 420

aaggggttt cttcctgttc gtgtttcggt gtccgcccga ctatcccatc cacccacctc 480

gggtcaaact gatgacaacg ggcaataaca cagtgaggtt taaccccaac ttctaccgca 540

atgggaaagt ctgcttgagt attctaggta catggactgg acctgcctgg agcccagccc 600

agagcatctc ctcagtgctc atctctatcc agtccctgat gactgagaac ccctatcaca 660

atgagcccgg ctttgaacag gagagacatc caggagacag caaaaactat aatgaatgta 720

tccggcacga gaccatcaga gttgcagtct gtgacatgat ggaaggaaag tgtccctgtc 780

ctgaacccct acgaggggtg atggagaagt cctttctgga gtattacgac ttctacgagg 840

tggcctgcaa agatcgcctg caccttcaag gccaaactat gcaggaccct tttggagaga 900

agcggggcca ctttgactac cagtccctct tgatgcgcct gggactgata cgtcagaaag 960

tgctggagag gctccataat gagaatgcag aaatggactc tgatagcagt tcatctggga 1020

cagagacaga ccttcatggg agcctgaggg tttagaccct gctcccatct ccccttcccc 1080

cactcaagag tcccagcaga atcccttccc cccaccccag ggatggagag gcactgtgta 1140

tctccctcca gactcgaagt catcctgcaa gatggcaaga accaagcaag ctccgatccc 1200

agggtgtggg agtgggggcc tgttcccggt ctgacctcct tggcactgga gcatctgggg 1260

cttcgttcat ccattcatcc cgtatcaggg ggccaaggta cctttacagg agcacctaga 1320

gcgagggcct ttggcaaaaa caaaacaacc aacacacctc tccacagggc cagctcctta 1380

gggataagtg gaagatggaa attgcaattc caagagggag tgtgcccaaa tgatttatgg 1440

ggatacctgg aagggagctt ggggtggggg ctgtctgtga cacttaagca gtctgggtgg 1500

ttgtctattt gtctgtcttc agtcttgaag cagggcttcc caatgccctt ttcctccctg 1560

ccttccttcc cccattattt cccacaggcc agcataattt tgtttttcct aatttatagt 1620

cactgttcta gacagaccaa agagaaggaa cagtggtgga gtctaggctg ctgatcagta 1680

agctttacct agcacctgag caccttttctc ccctcccctc tttcctcacc cttttctaga 1740

tgtaagacag aaagtaaatg tgactgggac ttaaccaagg tcttggtaaa gcctgcatgg 1800

caccgtaaga agctgaaaat actgtttgtt cccgcaatca ctgatttgaa aagttcccaa 1860

cacaggcagc tgctgtgtat atgggattag agccactaca tagaatagtc tcttacagat 1920

tttcataaat actagtcaca ataagggtat ttttcttggg ggtggagtaa gggggagact 1980

gatgctagtc cttgttgtat tttgttgggc tgtccttgtg tattttcacc ccagcctgta 2040

gtcctcctca cttcaacccc agggatttgt ggggagcaag ggtagccaat ggcagagggg 2100

gttggggctg ggactctgga ggctcctccc cttctttctc ttccttctgc ctcccccgtg 2160

cccccagctg ctcttgtcac tgtctctgat gggtatttgc ctggctttgt tgcttctcta 2220

tctgtattta gctgcagtga tcctttagct ggttggctca gaaaaaaaaa aatgtgcttt 2280

aggtgccctg taatcctggg catcaaggga atccatcctt ccccttttg atatgttctc 2340

cccgtacttc cagatttatt gttatggctc ccagtgggta ttggcgattc ttgtgatgca 2400

gggcctcagt cagtgtccag ccatgcataa gggagaggat agtgtgtacc tgccctgccc 2460

tctgctatga aggtctctgc cttgtggatc atgggactcc ccttggagga tctgtgcaaa 2520

ggggggctgg gcacaaagga gaatgtccta tttgggaggg caggaagcaa aggaactgga 2580

cagggattgg tgggcttggg gaacggaagt ttatcttgga taccccttgaa gaggctgggt 2640

ctcttcacat gaagatcgaa aagggaccct gcttccaatt tccctcttcc attcctcgag 2700

ctactccagg gctcagaaga atgctcttgg tctgtgggtc cagtgttgtc tgtcatccat 2760

ttaagtgttc ccactttcaa gtgacaatcc tctccttggc cctgccatag ggcagagcat 2820

gtctggcata gcagcctgac ttttatgccc taatcttgag ttgaggaaat atatgcacag 2880

gagtcaaaga gatgtcttta tatctgactg ttataaatga agttttgcgt tctgttgggt 2940

tccctttgg ctgccatcaa cgtcgtcttg cacgccagac gtgctgggga ttggagggct 3000

tggttcagca cgcattggtc ctttgtcacc ccctcctagg ccgttgtggg gcacgacgcc 3060

cgattgttgg acccccttctc tgtcgcggcg gcgtggcgca cgccttctcg agggccacgc 3120

cccgccttgg gcccgccccc ggggctcagg aggcctcctg gtgcgcgtct ccgggcgcct 3180

cccgcaccgg t 3191

<210> 55
<211> 385
<212> DNA
<213> Homo sapien

<400> 55
actgctaacc cctgccaggc ccagctgcca cacccttttct gggagaagca tggcctacag 60

aatgaagagg gggaccagga acccctgtgg gagaggctta gacctgaagc agtgcccact 120

ctggctcctc ctgccttggc tgactgggtt cctggaccat gtgcatttca ctgggccatg 180

ggatctacat ctccttgcat ccccagctgg tctgatccct gccagggccc cttccttcct 240

gctcatggtc ttcaggtggc ctgatcatgg aaagtaagga gttaggcatt accttctggg 300

agtgaaccct gactccatcc ccctattgcc accctaacca atcatgcaaa cttctccctc 360

cctggggtaa ttcaacagtt aaaag 385

<210> 56
<211> 1977
<212> DNA
<213> Homo sapien

<400> 56
ttcacccggc cctcgcctgt caccttcaca gagaaggggc tgcctgcagg aagaaagcac 60

ggcccacgcc cctccagtca catactgcct gtgggccctg gtgtagtgag gggggctgcc 120

gaggcactgc tgcactcacc aggcacgtgc tggtcggcaa acttgatgtt gatgatgtag 180

ttgcctggct ctgtggggca gtaggtcacc tggcgggaca ggtgcctggc tgctacaaac 240

catgcaatga gccatgcccc gccctggaca cccccgccca gcatctgggc ctccacgctt 300

gggaccgtgg gagcggccaa cagagctatg tctggagaca tatgataaac cacctcagcc 360

cccaccaagc cgccgcaccc gtagaccaga ccccaaggac cctggccacc atgggccaga 420

gagcattacc ttcatctctg gctctgctga gccggccctt gagtccccca cctgctgcct 480

gctctggcga ccctgggtgt gggagtggtg ccgggctgcc ttctgcttcc gcccgctgcc 540

gggattgcct ccagcgctgt ggaggcctgt gtgcggggat gcagcccctg cctgtctact 600

gaggactccc actgagggga ctgctgaagc caactggtgc caaggagcac aatggagtgc 660

cccccagccc tgatcgtgca ccccccagac cggcgggatg gccaggcggg ctgcaagtca 720

accatgggca gcagcttcag ctaccccgat gttaagctca aaggcatccc tgtgtatccc 780

tacccgagag gccacctccc cagcccctga tgcggactcc tgctgcaagg agccactggc 840

cgatccccca cccatgcgag cacagcctgc ccagcacctt tgccagtagt cctcgtggct 900

ccgaggagta ctattctttc catgagtcgg acctggacct gccggagatg ggcagtggct 960

ccatgtcgag ccgagaaatt gatgtgctca tcttcaagaa gctgacagag gctgttcagc 1020

gtacaccaga tcgatgagct ggccaagtgc acatcagaca ctgtgttcct ggagaagacc 1080

agtaagatct cggaccttat cagcagcatc acgcaggact accacctgga tgagcaggat 1140

gctgagggcc gcctggtacg cggcatcatt cgcattagta cccgaaagag ccgtgctcgc 1200

ccacagacct cggagggtcg ttcaactcgg gctgctgccc caaccgctgc tgcccctgac 1260

agtggccatg agaccatggt gggctcaggt ctcagccagg atgagctgac agtgcagatc 1320

tcccaggaga cgactgcaga tgccatcgcc cggaagctga ggccttatgg agctccaggg 1380

tacccagcaa gccatgactc atccttccag ggcaccgaca cagactcgtc gggggcaccc 1440

ttgctccagg tgtactgcta acccctgcca ggcccagctg ccacaccctt tctgggagaa 1500

gcatggccta cagaatgaag agggggacca ggaacccctg tgggagaggc ttagacctga 1560

agcagtgccc actctggctc ctcctgcctt ggctgactgg gttcctggac catgtgcatt 1620

tcactgggcc atgggatcta catctccttg catccccagc tggtctgatc cctgccaggg 1680

ccccttcctt cctgctcatg gtcttcaggt ggcctgatca tggaaagtaa ggagttaggc 1740

attaccttct gggagtgaac cctgactcca tccccctatt gccaccctaa ccaatcatgc 1800

aaacttctcc ctccctgggg taattcaaca gttaaaagaa gcttatctta aatgtattgt 1860

attgggggggt gggcagggcc cactctatgt tatgttaagg agttggttct ggttcttggc 1920

tgatgttctg tatcttaaca tgaccacagt ttgtaagtac ctcgctcgcg accacgc 1977

<210> 57
<211> 629
<212> DNA
<213> Homo sapien

<400> 57
tgggtgatct agtctgtcga gcgcgcatgt gatggatagc gtggtcgcgg gcccggaggt 60

tacatggtgg cctatacaaa agcactccat gtttctgcct gaataacgta cgttgatcgc 120

ttggttgcct gagcaaatgt ataagggagc tgcttggagg ggaaaggaac acaataaaac 180

gcctctggag gtatttcaaa gagtagtttc tcaaatctct ctgattcagg aagaggatga 240

cgagagggag agaacgtgga attacttaaa gtcatctaac tcattggtct tgttcaacaa 300

gaaagaattc tggtttgtgg ctgaatctga tttaacagct gccaatagct cattacttct 360

aagatgtata tctaattcaa aactagatgc tccccctcc ctctttttc cctaaaagct 420

gggaacttgt aagtaaatgg cctagtgaaa tagacactgg ctattttaga aaaagggtca 480

gaaggtagac caggaactat gtatcaagaa gaagtaagtg aggctcaagc ccaggccacc 540

taggctgtgt ttattcagtc cacccaatcc aatgtcaaag ccccaagtgg gccagtgtta 600

cctttatctt cccatctttg tatttctga 629

<210> 58
<211> 3535
<212> DNA
<213> Homo sapien

<400> 58
gtgataggca gctttccttc ttttcaacag tgatacctac gaaaatcaaa ataaatgcaa 60

gctgaggttt tgtgctcact gaaagggctg tcaaccccag aaggccgaca caaaaaaaat 120

ggtatgtgaa gatgcaccgt cttttcaaat ggcctgggag agtcaaatgg cctgggagag 180

ggggcctgcc cttctctgct gtgtcctttc ggcttcccag ttgagctccc aagaccagga 240

cccactgggg catataaaat ctctgctgta tcctttcggc ttcccagttg agctcccaag 300

accaggaccc actggggcat ataaaaaagt caaaaatcaa aatcaaacaa caagttctga 360

gttacttagg aaacagactt cgcatttcaa tcagagaggc cacagagcaa ggtctaaact 420

tctggcttct agacaaattc ctgatagaac atttaaatgt gggaagtggc ttccccaggt 480

cccatcccct gtttagggat agagttgata tcatttttat aggtgccatg tatgcctctg 540

cctgaatttt tttaattgac ttttgagctt ttgagattgc acgagggaga acaaggcctt 600

tgctgttgtg gataggaaag acttaaccta aaattaaacc agcaagaaag cattagtaaa 660

aatctaacaa tatgaagggc tcttatgagt catttttttc aaaagatgaa aactccagaa 720

acgcacagga acgaaatacc tcccagaaac atgaagcaat catcgaagac tcactggtaa 780

tatttttaaa aagtatacag atcaaagcaa aaagaagcca tgtgtaacaa agagaaatgt 840

gcaaatattt tttaaggcag tattaagtgc aagaggagta acatgaaata aacattcttt 900

cacatggcta ctgggaatat aaatttcgct ccagaaaggc cgtagcagtt tgacgatagg 960

tggcaaaacc ttaagattgt gtactggggc ccagaatttt tatttctagg aatgtatcct 1020

gaggaaatta tccgagatcc ccacaaactg caatgtttag gaattgtcct tatagcattg 1080

catacacaag aaaaacagag aaaagcctga tccctgtcag tggaaaaggg gttcaatgaa 1140

ttacggtgtg tctgcatgag gcttttatga cattaaaaat tgttgaacaa cggcctggca 1200

cagtggctca tgcctgtaat cccagaactt tgggaggcca aggtgggcgg ataacttgag 1260

gtcaggagtt tgagactagc ctggccaata tggtgaaacc ccgtctctac taaaaataca 1320

aaaattagct gggcgtggtg gcaggcgcct gtagtcccag ctgctcggga ggctgaggca 1380

ggagaatcgc ttgaacccgg gaggcggagg ttccagtaag ccaagatcgt gccactgcac 1440

tccagcctgg gtgacaagag tgagactccg tctcaaaata aaaaaaaaaa aaaaaagtaa 1500

aacaaaaata aagtctatgc ccattaagac gtcttctaat tcagttgtga ttgtctgctc 1560

ctacttaaaa aaatatttaa gcttgatgtt taattattcc ctttcagcaa atttggatca 1620

gaaaattaaa gtatgtgaca agatcaggtc accttgaatt tccacacaat ctcaagacac 1680

tgaatagcaa aaaagtaaca ttacacagta atgattagga tatttcctta gactttgctg 1740

gatctttggt cttaaggtaa catgtaaaag tagtgaagcc tttcctttca tggccctgtg 1800

caatgtaacg gttttctgcc tcctcttcag ctggaagcgt tagtggtagt atgggcacag 1860

aatatatgta cactggcgat gctgaccatg cctcccaggt accctggctc tgggttcctt 1920

gacctaggga acaagattgg atgaggcaga tctttgagcc catgtgacta tagaatttgc 1980

tgatgatata attttacaat aacaatggat aggaatttta cctctctttt tattagttta 2040

atattattta atattatgta cataagtgtt cactcgccta attaaaaaca ttgagtaaac 2100

caagttttta tatagactac ccttgccata tgatgctctt tttctctaat aatatgcagt 2160

ttaaatcctg aggaatcaat gcccagcatt tcaccacatc tgaactctgt gtgggcattc 2220

ttcactcgcc tacaaggggt aaacaaggct accagaactt gaatttgact tatagggagc 2280

tacccaggaa ggggaaagcc cttgggactt tttccaaaac aatcttctat ttgaactgtt 2340

catcagccaa agtagtccac tgaggtgaca aagctttcag aaatacaaag atgggaagat 2400

aaaggtaaca ctggcccact tggggctttg acattggatt gggtggactg aataaacaca 2460

gcctaggtgg cctgggcttg agcctcactt acttctcctt gatacatagt tcctggtcta 2520

ccttctgacc ctttttctaa aatagccagt gtctatttca ctaggccatt tacttacaag 2580

ttcccagctt ttagggaaaa aagagggagg ggggagcatc tagttttgaa ttagatatac 2640

atcttagaag taatgagcta ttggcagctg ttaaatcaga ttcagccaca aaccagaatt 2700

ctttcttgtt gaacaagacc aatgagttag atgactttaa taattccact tttctctccc 2760

tctcttctcc tcttcctgaa atcagagaga tgagaaacta ctctttgaaa tacctccaga 2820

ggcgttttat tgtgttcctt tcccctccaa gcagctccct ttatacaatt ttgctcaggc 2880

aaccaaggac agagtatcgg cagaaacatg gagtgctttt gtataggcca cctgtacata 2940

aaagtgtaat tatttattta attttcccat ttgtatcata ttaaagcttt gtacagtgtt 3000

ttaagttctg ttttaaaatt attttgtatt ttatttttat aacctagtaa taaaatattc 3060

attccgcatg caaaaaaaaa cacacacaca acccaaacaa ccaaaaacaa acagaaccaa 3120

aagatagaaa ccaaacacag caacagacag acacaagaga agacgaaaac aaccacacaa 3180

acagcacaca caaccgcagc ggagaaccaa caaagccaaa cgcaagacag cacaacaagg 3240

gacaaaacac acacgtagca caaagcagcc gcagaaccga acacacatct aaagacaggc 3300

gacaacaaca tagcaggccg tagcaccgac aaccactaaa ccataactat ccagcggaga 3360

tcagatcaca gaagccgaca cagaaagaaa agcatgtatg caacgcatga caccgcaggg 3420

cggataaaaa cactcgtagg cgggaccgcc gcagaacgat cgaaggctaa caaacaagct 3480

agtagatcac aagtacgtga ccacgccctc ttatctataa aggaggaaga gggag 3535

<210> 59
<211> 348
<212> DNA
<213> Homo sapien

<400> 59
cgaatgatgc tatatagggc catgggtgca ctgatgcatg gctcgagcgg ccgcagtgtg 60

atggatcgcg gcgaggtact tctgccatca ttttattcca ttccagcgct ctggcatgca 120

agataatcca tctctaaaat tcaagacttc caaattgaga tgaacgatta tgggcttggg 180

ttggggttta taaccaatcc gatcattgat catttgttcc cggctcttgg gatcactgct 240

aagcccaatg gctccttctc catcactgcc tcctacaact tccacatctt ccttctgttt 300

cttacgggtc tccaggtcct ttcgaatgtt ctcaaactct tcaatgtc 348

<210> 60
<211> 2497
<212> DNA
<213> Homo sapien

<400> 60
ttataccta tgctcgacca cttggagtgt tcaacctcgt tcaccggggt tttaatggtg 60

aaaccacgct tgctcgtccg tgtttcttgg aacgcttgcc ctgtaccata cctatgcgcc 120

cgtggtctat acggaagggt ctaaaacttt tctacggcta atgcggccca caaagttttt 180

tctttgtgtg tttttgtttg gatgcctcac agtggtccaa aatatattta tttttaaaaa 240

cagaaaatca gatcagtttt atagagtcaa attttcaaga gacaaaccag agtttggatt 300

ccagcttaga gtgcaagtga agcatcagta atctcatgca gaagatttgt ctctgcagga 360

ggtacattct gcttgactgc aagcactcaa ttctctaaat ctggtttgta cttctgccat 420

cattttattc cattccagcg ctctggcatg caagataatc catctctaaa attcaagact 480

tccaaattga gatgaacgat tattgggctt gggttggggt ttataaccaa tccgatcatt 540

gatcatttgt tcccggctct tgggatcact gctaagccca atggctcctt ctccatcact 600

gcctcctaca acttccacat cttccttctg tttcttacgg gtctccaggt cctttcgaat 660

gttctcaaac tcttcaatgt catcaagctt tagctctagg cgtgttggtt tccgtctcag 720

catactgaag tcaacactaa gggccaaacc cagtgaacta ttagcaattc acaggattat 780

cctccgcgtc caaacagatc ttaaatgtgt acgtcattca ttaaacaaac acttcttgtg 840

ttccagctac agagttaaaa ggtacagtca taatctcttt tcaagccggc ctagcccctt 900

cccggaacct cggctccccc ccaacgaaac tactgctaag ccaactggac tacacttccc 960

agactgcttg gagcctctct ctccgcagaa cctcgtcttc cgcgagcttt tcctggaggt 1020

tctaggaggg atgcccctca atgccacgac gccatttcct actacgactt ccatcatgct 1080

ccgcgccgct ccgggggcgg tgacccctct tggcccacgt cttgtcagtg acgcacttcc 1140

tgcaaccaaa actaaagcac ccgacgactt agttgctccg gtcgtgaaga aaccacacat 1200

ctattatgga agtttggaag agaaggagag ggagcgtctg gccaaaggag agtctgggat 1260

tttggggaaa gacggactta aagcagggat cgaagctgga aatattaata taacctctgg 1320

agaagtgttt gaaattgaag agcatatcag cgagcgacag gcagaagtat tggctgagtt 1380

tgagagaagg aagcgagccc ggcagatcaa tgtttccaca gatgactcag aggtcaaagc 1440

ttgccttaga gccttggggg aacccatcac acttttgga gagggtcctg ctgaaagaag 1500

agaaaggtta agaaatatcc tctcagttgt cggtactgat gccttgaaaa agaccaaaaa 1560

ggatgatgag aagtctaaaa agtccaaaga agagtatcag caaacctggt atcatgaagg 1620

accaaatagc ttgaaggtgg caagactatg gattgctaat tattcgttgc ccagggcaat 1680

gaaacgcttg gaagaggccc gactccataa ggagattcct gagacaacaa ggacctccca 1740

gatgcaagag ctgcacaagt ctctccggtc tttgaataat ttttgcagtc agattgggga 1800

tgatcggcct atctcctact gtcactttag tcccaattcc aagatgctgg ccacagcttg 1860

ttgtgatgaa ccagtggcag atattgaagg ccatacagtg cgtgtggcgc gggtaatgtg 1920

gcatccttca ggacgtttcc tgggcaccac ctgctatgac cgttcatggc gcttatggga 1980

tttggaggct caagaggaga tcctgcatca ggaaggccat agcatgggtg tgtatgacat 2040

tgccttccat caagatggct ctttggctgg cactgggggа ctggatgcat ttggtcgagt 2100

ttgggaccta cgcacaggac gttgtatcat gttcttagaa ggccacctga aagaaatcta 2160

tggaataaat ttctccccca atggctatca cattgcaacc ggcagtggtg acaacacctg 2220

caaagtgtgg gacctccgac agcggcgttg cgtctacacc atccctgctc atcagaactt 2280

agtgactggt gtcaagtttg agcctatcca tgggaacttc ttgcttactg gtgcctatga 2340

taacacagcc aagatctgga cgcacccagg ctggtccccg ctgaagactc tggctggcca 2400

cgaaggcaaa gtgatgggcc tagatatttc ttccgatggg cagctcatag ccacttgctc 2460

atatgacagg accttcaagc tgtggatggc tgaatag 2497

<210> 61
<211> 604
<212> DNA
<213> Homo sapien

<400> 61
catctgatca cgtcgcccgc cctagtctgc ttttgtgaat ctccactttg ttcaaccccc 60

acccgccgtc tctccctccc tgcgccttcc ctctaacctt aaagacccag ctacctctac 120

gcaaatggta gctacctctg cgcaggtgga agttgagttc ggttcacgcg ggaccctctt 180

ccctgtggca agctgctgaa ggagacctgc tgttactgct ttagctgctg ttctgcttaa 240

tctcagacag ctggaaggta gtattccgcc ctcgcccagc cacccgagta aatgagaatg 300

atgactgact gagaccctga tggtggtggc ttaggcttaa tctcagacag ctggaaggta 360

gtattccgcc ctcacccagc cacccgagta aatgagaatg atgactgact gagaccctga 420

tggtggtggc ttaggcaaga tgaacgtctg aaatggaggc ccagggctgc cacgatggct 480

ctgtagtcat cagggaggga gctcctttta tcttgttacc tactcccctc ttgtgcccct 540

ttcttccctt gatctaagaa ggttgctccg ggatctcgcc atctggcctg cgttcaatca 600

gcag 604

<210> 62
<211> 4733
<212> DNA
<213> Homo sapien

<400> 62
atggagagaa accagaacag aaaagtggaa aattccaaaa accagagcgt ctcttctcct 60

ccaaaggatc acagctcctc accagcaagg gaacaaaact ggatggagaa tgagtttgat 120

gaattgacag aagtaggctt cagaaggtca gtaataacaa actcctccga gctaaaggag 180

catgctctaa cccatcacaa ggaagctaaa aaccttgaaa aaagtgttaa cactcctgtg 240

gacgttttgt ctttgttttt caccccagtc cacacctctg ctgttagtga ggggctcctg 300

gtcatcacag ccacatgtgg actcccctgc tctgtacctc ccagcctttc tgccaacagc 360

actggcctca cagtgctgtc ctggggattg agtgagatag ctataatgtg gagtcgctgc 420

tgcctgctaa ctcgcccaga cgttcttcga gtgtcacatt tgaagacgag cactgaggat 480

gaggaaccaa ctgaagaata tgaaaatgtt ggaaatgcag catctaagtg gccaaaagtg 540

gaggatccta tccctgaatc taaggttggt gacacatgtg tttgggatag caaggtagag 600

aatcaacaga aaaagcctgt ggaaaacagg atgaaggagg acaaaagcag catcagggaa 660

gcaatcagca aagccaagag tacagcaaat ataaagacag aacaggaagg tgaggcatct 720

gagaagagct tgcatctgag cccacagcat atcacacacc agactatgcc tataggacag 780

agaggcagtg agcaaggcaa acgtgtggag aacattaatg gaacctccta ccctagtcta 840

cagcagaaaa ccaatgctgt taagaaatta cataaatgtg atgaatgtgg gaaatccttc 900

aaatataatt cccgccttgt tcaacataaa attatgcaca ctggggaaaa gcgctatgaa 960

tgtgatgact gtggagggac tttccggagc agctcgagcc ttcgggtcca caaacggatc 1020

cacactgggg agaagccgta caagtgtgag gaatgtggga aagcctacat gtcctactcc 1080

agccttataa accacaaaag cacccattct ggggagaaga actgtaaatg tgatgaatgt 1140

ggaaaatcct tcaattatag ctctgttctg gaccagcata aaaggatcca cactggggag 1200

aagccctatg aatgtggtga gtgtgggaag gccttcagga acagctctgg gctcagagtc 1260

cacaaaagga tccacacggg ggagaagccc tatgaatgcg acatctgtgg gaaaaccttc 1320

agtaacagct ctggccttag ggtccataaa aggatccaca caggtgagaa accttacgaa 1380

tgtgatgagt gtgggaaggc cttcattact tgtagaacac ttctcaacca taaaagcatc 1440

cactttggag ataaacccta taaatgtgat gagtgtgaga aatcttttaa ttatagctct 1500

cttctcattc agcataaagt catccacact ggagagaaac cttatgaatg tgatgaatgt 1560

gggaaggctt tcaggaacag ctcaggcctc atagtgcata aaaggatcca cacaggagag 1620

aaaccttaca agtgtgatgt ctgtggcaaa gcattcagct atagctcagg cctcgcagtc 1680

cataaaagca ttcaccctgg gaagaaagcc catgaatgta aggagtgtgg gaaatccttt 1740

agttataact cactacttct tcaacacaga actattcata ccggagagag accttatgta 1800

tgtgatgtgt gtgggaaaac gttcagaaac aatgcaggcc tcaaagtcca caggaggctc 1860

catactgggg aaaaaccata taagtgtgat gtgtgtggga aagcctatat ctcacgctct 1920

agccttaaaa atcacaaagg aatccacctt ggggagaagc cctataaatg tagctattgt 1980

gagaaatcct tcaactacag ctctgccctt gaacagcata aaaggattca taccagggaa 2040

aaaccctttg ggtgtgatga gtgtggtaaa gctttcagaa ataattctgg ccttaaagta 2100

cataaacgaa tccacactgg ggaacgacct tacaaatgtg aagaatgtgg gaaagcatac 2160

atctctctct cgagccttat aaatcataaa agtgtacacc ctggggagaa gcccctttaag 2220

tgtgacgagt gtgagaaggc cttcatcaca taccgaaccc ttacaaacca caaaaaagtt 2280

catcttgggg agaagcccta caaatgtgat gtgtgtgaga aatcttttaa ttacacatcg 2340

ctcctttctc agcacagaag ggtccacact agagagaaac cctatgaatg tgacaggtgt 2400

gagaaggtct tcagaaacaa ctcaagcctt aaagttcata aaagaatcca tactggggag 2460

aggccctatg aatgtgatgt gtgtggaaaa gcctacatct cacactcaag ccttattaac 2520

cataagagta cccaccctgg caagacaccc catacatgtg atgaatgtgg aaaagctttt 2580

ttctcaagca gaactcttat aagccataaa agagtccatc ttggggagaa acccttcaag 2640

tgtgttgagt gtgggaaatc tttcagttac agctctctcc tttctcagca caagaggatc 2700

cacacagggg agaaacccta tgtgtgtgat aggtgtggga aggccttcag gaacagctca 2760

ggcctcacag tgcataaaag gatccacaca ggtgagaaac cctatgaatg tgatgagtgt 2820

gggaaggcat acatctcaca ctcaagtctt atcaatcata aaagtgtcca ccaggggaag 2880

cagccctata attgtgagtg tgggaaatcc ttcaattata gatcagtcct tgaccagcac 2940

aaaaggatcc acactggaaa gaagccatac cgatgtaatg agtgtgcaca tatacccaac 3000

gccaccgcgg acctcatgaa agtggaccat gaagaggagc cccagctctc cgagccctac 3060

ctttctaaac aaaagaagct catggccaag atcttggagc atgatgatgt gagctacctg 3120

aagaagatcc tcggggaact ggccatggtg ctggaccaga ttgaggcgga gctggagaag 3180

aggaagctgg agaacgaggc actttcccag tggaaagaat ttgatgccat ttccgcacac 3240

ccaaattcct gttggcagga taatcagcaa gtaatacaga gagctgcttc cttcaacagc 3300

tgcattcagc acaaaaccga gggccgatta tgtggagaaa tatccttcag agaagcatgt 3360

gatctttgtt ctgccttttg ccctttgaag catatgatct ttgttcctac tccctgttcg 3420

tacactccct ccccttttga aatccttaat aaaaacctgc tggttttacg gctcaggaag 3480

cttctgctga ttgaacgcag gccagatggc gagatccgga gcaaccttct tagatcaagg 3540

gaagaaaggg gcacaagagg ggagtaggta acaagataaa aggagctccc tccctgatga 3600

ctacagagcc atcgtggcag ccctgggcct ccatttcaga cgttcatctt gcctaagcca 3660

ccaccatcag ggtctcagtc agtcatcatt ctcatttact cgggtggctg ggtgagggcg 3720

gaatactacc ttccagctgt ctgagattaa gcctaagcca ccaccatcag ggtctcagtc 3780

agtcatcatt ctcatttact cgggtggctg ggcgagggcg gaatactacc ttccagctgt 3840

ctgagattaa gcagaacagc agctaaagca gtaacagcag gtctccttca gcagcttgcc 3900

acagggaaga gggtcccgcg tgaaccgaac tcaacttcca cctgcgcaga ggtagctacc 3960

atttgcgtag aggtagctgg gtctttaagg ttagagggaa ggcgcaggga gggagagacg 4020

gcgggtgggg gttgaacaaa gtggagattc acaaaagcag actagggcgg gcgacgtgat 4080

cagatgacct gtgcgggcgg cagcctcctg ccctcctccc cttcgtgcgc cggctggagc 4140

gaagagttct tttgacagcc gtgagcttcc ccgccaggaa cttactgggg ctgcatcacc 4200

ctagaaacgt ggctttgggc tgtggaaacg ctgcctctgt ggaagtctct cctcgcgggg 4260

gtggacgggt cgctgcgcgc ccagcgttct tctgcggttc tcacagcccg cccgccgccg 4320

ccgcctcggg gacccctttgc ggggaggcct caggtcagcg ccgcccccttg cgatggcggg 4380

agagcagagc ggccccaggg cctctgagct ccggccccgg cacgtcccgc cctctcctcc 4440

gcttgcggga gccggggcgc ctcggaggga aaccttcccg gacacaggcc gggagaagag 4500

gggcccgtgg cgccatcggc gcagcgttcc tggtttcggc catgtttttt ttttttttt 4560

tgagacggag tttcgctctt tttgcccagg ctggagggca atggcgccgt ctcggctccc 4620

tgcagccgcc gcctcccggg ttgaagggac tctccggcct caggccgcgg tgttcttgta 4680

aattagaacc ggcctatgcg ccaagcgcgg gtctcgcggc tgcggagaga ggc 4733

<210> 63
<211> 577
<212> DNA
<213> Homo sapien

<400> 63
aaaaacaaac aaaaaaacca gttatttact tggctataac aatgtctgta aatggtaaag 60

acagaaaaca acctaagttg ttcattgggg gactagataa atttaaattg gttacattcc 120

atacaatgtg gaagataaag tagctgtgag gaacgatgcg cctcctccct ttacagatgt 180

ggaatgaccg ccaagataca gccaggaaaa aagcaagaag agaacagttg gtgatctgct 240

gtttcagttc aaaataacca attttaaaaa aatgtctcta gaaggctaca taagaggctg 300

gtgagacagt ggtcctgtgg gagggaatgt gggtaacagg caacaggatg ggagggatgg 360

tcaccataca tcttttggca cactctgcat tttgaaacat atgattggtt ttccctagtc 420

taacaaaata actttaggtc caactgactg tagtaaattg gtcactagca aaaaaaaaaa 480

aaaaaacaaa aaaagcgtgg ggggaacccg gggccaacgc gggtccccgg tggggaatgg 540

gtttcccggc tccaatttcc ccccattttg cgcacaa 577

<210> 64
<211> 744
<212> DNA
<213> Homo sapien

<400> 64
aaaaacaaac aaaaaaacca gttatttact tggctataac aatgtctgta aatggtaaag 60

acagaaaaca acctaagttg ttcattgggg gactagataa atttaaattg gttacattcc 120

atacaatgtg gaagataaag tagctgtgag gaacgatgcg cctcctccct tacagatgtg 180

gaatgaccgc caagatacac agccaggaaa aaagcaagaa gagaacagtt tgtgatctgc 240

tgtttcagtt caaaataaca atttaaaaaa atgtctctag aaggctacat aagaggctgg 300

tgagacagtt gctctgtggg agggaaagtg ggtaacaggg aacaggatgg gagggattgt 360

caccatacat ctttttgcac actctgcatt ttgaaacata tgattgtatt tcctagtcta 420

acaaaataac attaggtcca actgactgta gtaaattgtc actagaatct ttgtcttgga 480

gtcatttgat tttatttttc acagaagggc atatggttag atagcctaaa acagcagggg 540

agggagcagt gtgtgctagg agctgccaca tgccaggcac ctgatagata ccagctctat 600

ctcacgctcc ccctgtccct tgaggtaggt gtttctgttt ctacttacaa gaaaatagac 660

taaggcgtaa agtaacacaa tcacggtcac acatgtggca agcagcacag cctggatttg 720

aagctaggca agtctgactc taag 744

<210> 65
<211> 318
<212> DNA
<213> Homo sapien

<400> 65
taagtgtttt aaagaaacta gagcagcaga gaagagagac tcagaagccc gtggcacggc 60

tttctcggcg tcccctccag cgaggggggtc tccattgctg cagttgccgg ttttgtccaa 120

ccaggtcagg aggctgcccg gccccctccc cactctcaaa gttgcttgtt aaacacagag 180

tcgtaatttg tggctaaata taactagtgt gttctcacgg aaagtataat tcagggtgct 240

cattgtatga ggttatcaac aagacccatc tgggttaaat taaagtgatt ttcataaagg 300

gcagaagcgc ccctttac 318

<210> 66
<211> 2505
<212> DNA
<213> Homo sapien

<400> 66
ctgcgccccc aataatcgtg tgttgggtat ctctccattg ttccctgggc cattagtcaa 60

agggcaaatt agaaacaatt tcttgacgac aaaaattgac tcttaattct tcccaccaga 120

ggcgcagggg gacaagtcca gaggccaggc ctgaggctgg ggctggcggc aggaggggta 180

ggcggggtag aggtggggta gaggtggggg atctaagatg gttcacagca cagagacctc 240

cctcaactgg agggcagcag cagctcaccc ccaccccaga atcagtgttc cagaggacag 300

ctgggggcag ggggttagga acaatcccac tccctcctgg aagaggtcct gacccccac 360

cccacccatc tgcctgtacg ggtccctctg ctatcgactg gggcgagttc ttttcatgag 420

ggccttgcct ggtgcctgga gggaacaggg tgtgggggaa aggtctgtgt gttccccct 480

cacctccct gctcagcggt gtggcctctg gctctgggga ggatcgggaa gccccagggt 540

ccgtgccttg gagtgggggg acagcctttt ctctgccaca gcttctgcct ttgggggctt 600

acccttccta ggagcagaac tgttgtggag gggaagagga cggtgaaggt tccgaaggga 660

aggggggctgc ccccactgaa aacgaagctt ccagtcacag ccccttcatt atttatcagg 720

acccagggga tgaggtggca ggggaggggg ctgcatggag ggagtgccct caccctcgtc 780

cccagcgcct gccccctccc gaccaggcct gggctgaggc ccagggtaag ggggctgagg 840

caggccacag aggagcaaga cttgtcaggg gccagacctg ggtaggagga ttgtcctcca 900

ggcacacacg gcccccagcc ccccagcctg tcgaactggg ctctcccaga aggtccccgg 960

ctccagccca agcagggagc caggttgggg gtgtgggaag gcagaagtcc cagggatcct 1020

gggggaggct caggttgtac ctgcaggcca cagtctcctc gacagacctc ggacgaggtt 1080

gtaagtgttt taaagaaact agagcagcag agaagagaga ctcagaagcc cgtggcacgg 1140

ctttctcggc gtcccctcca gcgagggggt ctccattgct gcagttgccg gttttgtcca 1200

accaggtcag gaggctgccc ggcccccctcc ccactctcaa agttgcttgt taaacacaga 1260

gtcgtaattt gtggctaaat ataactagtg tgttctcacg gaaagtataa ttcagggtgc 1320

tcattgtatg aggttatcaa caagacccat ctgggttaaa ttaaagtgat tttcataaag 1380

ggcagaagcg ccccttttgcc tggtttcccc tgccttttta ttagtgacag tgttattgtt 1440

gcaattatta gaggggatca aggaaggtgg gagttgccag ggacccccag gctgaggagg 1500

gagcctccca gccccccctcc actcaccatt gtcctgctgc ccctccagcc ccctggtcac 1560

tgtcacctct gcggctggag ggctggagaa aggcccagga gcccagggaa cctccaccat 1620

gccaggctgc cttccctaaa ggccggctcc ttccccccag gcagccagtc tgggggtgga 1680

cctggagggg cagggttctc tgggaatctc tgggccaagg gatgcccttt ggggatctgc 1740

tggaataaag agtaatcggt tttccttgca ggaggttgac ccgaaacagc cctccgtctt 1800

cctgcccatg aggatgagga caggaagcca gacccgggag agaaagccca gatcagacag 1860

atgggggttg ggagggccta gtggctgaaa ctgccctgct tggggacaat gctcaggcag 1920

gaccaatcca aacgagggca ctgcttatgt ggtccttccg gagcatcctc caaggcagag 1980

cctggcttcg ggccctccct gcaaggacac ctgcctccac tccctgccag accaggcctg 2040

gctgcagagg cggaggggcg gcagggctgc ctgtctccac tgtttggcgg tgtttacacg 2100

tgtcgctgtc agggtgtctg gacagtgggg gtcgggggcc tcccgggccc agacttcccc 2160

acgctcccca agagcagcag atgtaggtct cccggctccc aaacctctga gatgctgctc 2220

tgtctgaaat catcatataa ataatcattt tttaatacaa gggggtgggg gtgtcaggca 2280

agttccatga aattgtgaat tagccgctgc tccaaataaa tgcctgcttc ggcccgagag 2340

cagccgcgct gcgcccgctg cgcccccaat aatcgtgtgc tgggtatctc tccatttgtt 2400

ccctgggcca ttagtccaaa gggcacaatt agaaacaatt tcttgacgac aaaaattgac 2460

tcttaattct tcccaccaga ggcgcagggg ggacaagtcc cagag 2505

<210> 67
<211> 247
<212> DNA
<213> Homo sapien

<400> 67
actgagcagc tacggaatgc aaggcactgt aggaagtagg gtgagtatac tccccacaag 60

ggctcagggt caggcagggg aggtgagata aaaacccaca gccatacact agctggcctg 120

tcctgagggt tgtgaggcac aaaatgctag gagactagag aagtaagaaa tgtcttgaca 180

tgaggagaaa tcaaggaaag catggtttta gaacatgtgt gggatgtttt ttgtatcgag 240

actgaag 247

<210> 68
<211> 2458
<212> DNA
<213> Homo sapien

<400> 68
ttgcgctgtt ttgggtgatc ttgtggtgcc gtcgcgcgcg tcatgcttgt tttctcagcg 60

gtgctatctg tgcgtcggta tgtgtgtctg ctgcacgggc tggtgcgacg cgtggtttgt 120

gctcgtgatg aatgggctgc ggcgtttgtg cctatggggc ctgttcaagt atgggggcct 180

ttgtgggcct tcggggattt ttgggggggcg cgcccgttgc ttgtggttgg acggccgctc 240

tgaggatgtc aggtagcccg ttactggtgt cggaagtttg tgcaagattg agataaatga 300

ttagtaatta caggaaaact aacttgtaaa aatcttaaag acattgaatg ggttaatgta 360

ctgagcagct acggaatgca aggcactgta ggagtagggt gagtatactc cccacaaggg 420

ctcagggtca ggcaggggag gtgagataaa aacccacagc catacactag ctggcctgtc 480

ctgagggttg tgaggcacaa aatgctagga gactagagaa gtaagaaatg tcttgacatg 540

aggagaaatc aaggaaagca tggttttaga acatgtgtgg gatgtttttt gtatcgagac 600

tgaagaggct ttttaaagtg gagggaaggc aaactgaggc atagagatgc caataccagg 660

tcttgtcagg aagaacagag tccaatttgg ctgcaggata gggcatatgt aggggagggg 720

ataagactgg catgggggca gaggggact tgaatgtcag gtgacagagt caaagcttgc 780

actcaagtag gtgattgata ggtgatcttg ggggcttttg agcaggaaga gtgataaatc 840

aacattggct taggacaatc cctttggctt tggcgaaatt gtgactcagg taagggacag 900

ttgggggaga caaggatggg aagatgcatg ataaatgggc cctgagcttg gcaggcaaca 960

ggaaggcagt gcaatgtttt gtagaaatag acttagcaga gcaggcgccc agcattcagc 1020

gagaccacag attcttagcc actgtatgac ttgaaggtga tactaactcc cctgaacctg 1080

tctcatctac aaaatgggca tagtagcacc tatgtcatag ggttgcttta tggcaggtaa 1140

ttattcttgt aaagcattta gcacagctct gatgacacat ggggaatgct cagtaaatat 1200

taatgtcact aattccagtc aattccatta actggaatta attagttgga gttaggaagg 1260

aaacaggtgt cagctgggtc gactgaagga agcagagatc tcaggaagca gtacatggtg 1320

gtacagtggg agcatggcat ctgtgggggа gactgtctgc agtgggatcc cacctctatt 1380

gctggccagc tgtgatcttg tacaggatac ttaacccctt tctagctcag tttctttagc 1440

tgtaaagcca gcgcaatgtt accatggagc tggtatgtga aagcaatgca tgttaagtgt 1500

gtagtatgtg cctgatacat tgttagtgcc cagtaaatgt tagttatcag gtgaggggtg 1560

ggatgtgggg ttgagtagat ttgcctcgtc cactcataga ggcaggatgt gcaaggaggg 1620

gtggagacat gcagcaaggg attgggcatg aaggttttct ccaggcccgg ggagaacact 1680

cggcctggag atattcagga gtccctggcc cagaggagat tacaactctg cagggatatg 1740

agccacgagt ggagtgtaga cataggccta aggaagcaat cagacacttc cagaatagag 1800

gacatttgct aagacagctg gcctggatgc ttcaaaaagt atcatgggaa aatatgtcgg 1860

ggttgttgta gactgaaagt gactagagtt gaaaccaaaa gaatgtatga accttgatgg 1920

gacctggttg ttgtcgttgt ttttaaagac aaacattctg gagacaattg ttgaatactg 1980

attaagggct ctgtattaga ggctataaag aagttactgt ttatattttt aggcatgatc 2040

atcttaatgt aatcatgtag aacggtgtct ttgtgctcag gaaatctatg tggtaatact 2100

caggggtcga aatgtcattc tatgtacaca tgcatggggt ggagagaata gtaaatctgg 2160

caaaatgtta acaactattc aatccaagtg gagagtatgt atctgagtat tcattgtgct 2220

agcctttcaa ctttttctgg atgtgtggaa tttttcaata taaaaaaata agaaaaaaaa 2280

gaggcctatg gaaaaactgg gaaccagtat gcaacaggaa aaggaattct ttgaaggaaa 2340

aacaggatcc acctaccaag agagcaattt caagggagag attaagcaca caatgaaggc 2400

tgcaaataga tcatcatatt tggtcactgg gagattgtgg cagcttggaa atagtccg 2458

<210> 69
<211> 894
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature

<222> (341)..(341)
<223> a, c, g or t

<400> 69
gatgataatc atataggcca atggtgactc tagatctgct cgagcgggcg cagtgtgatg 60

gatcgtggtc gcggcgaggt actccgctac tcttttactt cttttgtaaa gtattgactc 120

tggaaggcta cagtatacaa agtctcaaca tgtttttttaa aagaaataag gagcaagcga 180

ctgccctgct agaaatcaca aaccgatttt tgtagaatat tttgtgcccc aggcattaat 240

ttcactgact ccagaacctg cagttcagag aatgatttct tatgatgata aaaatcgaat 300

gggatcagac gatggtttgc atttttttta attaacttgg naataggaca cctcaagttt 360

gagatttcat tttcttttag aacacagtca caagattaat ctggtgaatc cttttgtcac 420

aggttctcgg tgtgtgtgtg cgcgtctccg tgtgtgtgtg tgtgcatgtg tgtaaaactg 480

gtcacattta attgcttttt gggaccattg aatagttggg aagtaagaat tttttaattg 540

gcattaaaac ggtcctcaac tgttaaatta accaaatttg acctgtcttt aaaaaaaggc 600

ttatttgtat gattttgggc taactccccg gggaccatat taaatgacaa aaatgctcct 660

ttgggtgaca caccctacaa agtatttgct gttacgaaca taaacgccca cattcttaat 720

atctaatatt tttgaccagt gatgtcttat gctgtcatct gaacctagag aagcagtgtc 780

agaggaaacc tggtgtccat gtgtcttagc aaagggtact gatcgagggt ctgtgacaaa 840

gatggcccga gagctggggt atcatggcat agtgttctgg gtaatgttcg ccac 894

<210> 70
<211> 1335
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (278)..(278)
<223> a, c, g or t

<400> 70
cgtggtcgcg gcgaggtact ccgctactct tttacttctt ttgtaaagta ttgactctgg 60

aaggctacag tatacaaagt ctcaacatgt tttttaaaag aaataaggag caagcgactg 120

ccctgctaga aatcacaaac cgatttttgt agaatatttt gtgccccagg cattaatttc 180

actgactcca gaacctgcag ttcagagaat gatttcttat gatgataaaa atcgaatggg 240

atcagacgat ggtttgcatt ttttttaatt aacttggnaa taggacacct caagtttgag 300

atttcatttt cttttagaac acagtcacaa gattaatctg gtgaatcctt ttgtcacagt 360

tctcgtgtgt gtgtgcgcgt ctccgtgtgt gtgtgtgtgc atgtgtgtaa aactggtcac 420

atttaattgc tttttggacc attgaatagt tgggaagtaa gaatttttta attggcatga 480

gacggttcct caactgttaa attaaccaac tttgacctgt ctttagaaaa aggcttattt 540

gtatgatttt gggctaactc cccggggacc atattaaatg acaaaaatgc tcctttgggt 600

gacacaccct acaaagtatt tgctgttacg aacataaacg cccacattct taatatctaa 660

tatttttgac cagtgatgtt ttatgctgtc atctgaaccc tagagaagca gtgtcagagg 720

aaaccttggt gtcacatgtg tcttagcaaa agggttacca tgatcgaggg tcatgtgacc 780

aaaagatgct ccagagaagc ttgagaattt gtttcaagtt gggaggaggg ttggagatac 840

aaaaatcact ctgctctaca ggactcttca gctgtctatg caagaaattc cgttttctct 900

ttcagcacct ggaaagacac agcagcccac tgaggcgata ggtgattcac taagcacaag 960

aggaatgttt tctaagcaag gcgtcccttg cctctcaaac aaatgccctc caagtttgtt 1020

agggtttcta ttcctgcaac ttgtggtatc aaaaccactt cctgaaattg tcaaagcact 1080

gccaaaataa atgtttttcc cccttctaag aaaaaaaaaa tgacagtgct catatttgac 1140

acttgtgtat tggactctct tttgaatgaa taaaaaggaa aaggggtttg gtgtaattcc 1200

tgatggggtg cgtgttgttt ttcatgccat ggtttgtgaa ttttaattgt ggtttcccat 1260

ttcgttgttg taactgggca gaaattaaaa aagaaaaatc aataaaaata caaagaaatg 1320

gttaaaaaaa aaaaa 1335

<210> 71
<211> 137
<212> DNA
<213> Homo sapien

<400> 71
cgggcaggta cgcacaagca ggcccagctg ggagacgcta ggataactaa gacctggtca 60

gcgccctgag gagtcttgtc tggataaagg gagacacaca ctagcttgga cttatgcgtg 120

cagcggagct gtgtaga 137

<210> 72
<211> 694
<212> DNA
<213> Homo sapien

<400> 72
ccgcccgggc aggtaccaac cccagcacac cccaacagcc tttcctcggc ccctcctcag 60

gcctcctaat tactctttct cagcctggag tgtggggccg ttaccgtcct cttccccctt 120

ctccttccat actgcactta accttgctgg aagacttaat gatggagatt tagggcaatc 180

tgtggctgct tggacccttc cctggaccaa aggaacttaa aacccaaacc tgacactgga 240

atgaaatcca agtttttaaa tatcaccttt caatcactca cagatctcac tctatcttaa 300

aatactcagc ctcactcctt aatgagtgct tgctgaaggg agaaaattcc attttaaaaa 360

cgtattcact ttactgatta ctgtgcaatt tgaattaagt cacgattctt tagtaaatgg 420

aggtgagaat ctcagattca aattgtcaga gaccatgatt tagaagtcta ccaaacaccc 480

agtttccttc cactgtttta gggtaacagg aaaacatgag attggggtgg tgtccgctat 540

taaatggaac cacacatcat gaaattcaat tctcatgtta agacattctg tattgtggga 600

tgtcaaaagt atttcccaaa ctttcgtttg acctgcagag ttggagatgg cttacctccc 660

tataacttca agtttgtttc acaaagcttg gcgt 694

<210> 73
<211> 8095
<212> DNA
<213> Homo sapien

<400> 73
ttttttttttt gccacctaga gatgataatt tattgtttta ccatgactca gaagagaaac 60

aacataaaga gaatatttca aatccccaca atttccttcc tcaacctcac tactcttaac 120

atttctttat cagacgccac tggcttccta aaatggacca ctgactatgt atgtgtacac 180

atttcattat gctgcctttt ctcttatgat taaaacttta gccctcattc gaggtttcca 240

atggttactt ttagtggagg agttccctag cttttaaaaa accacttttc ctctaagatt 300

ccattattta ttgaaagaag tctttctaga aatgttaagg aggattttaa atgaacacat 360

tcaattaaaa aaaaaatcac gtattgaaca tctaccaagc atctggactc ttcggaacct 420

agtaaaatga aaaaatccag ttttaacaac agtaacttca ttctgcgggt atacagagac 480

aagcacgttt cttcttttgg tctaatttat tctaaacgaa gaagctggga actgacaaaa 540

caggacaggt tgtttttaat ccagtctaca aataaacaag acaatgcctg agttagccct 600

ctatatagat ttaggcttat gctgacctcg ttgtaaaatc tgtatttaac taaaagttaa 660

taaaaataca tatgttcatt ttaaaataat tactgatttt gcttggctat cccacccctt 720

acccccaaac tcatatattt ttaggacaag attttcctgc ataaccacaa cctgtctcct 780

cccaccccac ccccatcata gatgttttca aataagaacc cctgcgatca gcagaagcat 840

ctctaatcta acatgctttg tccttgctag ggcaggctaa aagctttaaa aagcaaccgg 900

atgctcttct ctggttgagg tgaggggaag gcgctcgggt accaacccca gcacacccca 960

acagcctttc ctcggcccct cctcaggcct cctaattact ctttctcagc ctggagtgtg 1020

gggccgttac cgtcctcttc ccccttctcc ttccatactg cacttaacct tgctggaaga 1080

cttaatgatg gagatttagg gcaactgtgg ctgcttggac ccttccctgg accaaaggaa 1140

cttaaaaccc aaacctgaca ctggaatgaa atccaagttt ttaaatatca cctttcaatc 1200

actcacagat ctcactctat cttaaaatac tcagcctcac tccttaatga gtgcttgctg 1260

aagggagaaa attccatttt aaaaacgtat tcactttact gattactgtg caatttgaat 1320

taagtcacga ttctttagta aatggaggtg agaatctcag attcaaattg tcagagacca 1380

tgatttagaa gtctaccaaa cacccagttt ccttccactg ttttagggta acaggaaaac 1440

atgagattgg ggtggtgtcc gctattaaat ggaaccacac atcatgaaat tcaattctca 1500

tgttaagaca ttctgtattg tgggatgtca aaagtatttc ccaaactttc gtttgacctg 1560

cagagttgga gatggcttac ctccctataa cttcaagttt gtttcacaaa gctttgaaaa 1620

gtaaaacaga taatttcatt ttcagataat aaaaaatctg aatagcaaaa taattgcttt 1680

taaatgtagt gtgtccactc taaaaaaaaa aaccctaaat ctatgttaga aaaacttttc 1740

aaataatgcc ttttattaaa ttctccagta gtagttgaaa taaaaatcta ccctaatttc 1800

tatgaaatga tctatttata tcactgactt ttctttttct ctgattctat atttcattta 1860

acaatctgca gactttcacc ccatttccca gatgggaaaa ccctagcccc ctcgtatctc 1920

tgagaagttg ctcagagtag gacacagaga aatatggccc ccaccctggg aagtactgct 1980

gtcactgttt aagtgtattt cagttctgtt actccaattc atacacacag tcttccatga 2040

ggatggtagg atgaacctgg ttagctggct ttggataagt agatcagcat gactacctgg 2100

aataaaagtg actgactcta ggataaaaat taaaaaaaga ttctttcaca gcaacgagtc 2160

tttgcaaaac ctctctccta ataatcacaa accaggggaa gaaaagtggg agcagggaac 2220

acaggaacac agccaaaggg aatattgcaa aatgcttccc gagcttcatc agacagactt 2280

cttgcaatgc cacgactgga tgcatctgca cacaattccg ggaaatgccc accttgctgt 2340

tctcctatcc ccaattttct ttctttcttt ctttcttttt ttctttttttt tgagacagag 2400

tctctgttac ccaggctgga gtgcagtggt gcaatctcgg ctcactggca acctctgcct 2460

cccagattta ggtgattctc ctgcctcagc ctcctgagta gctgggacta caggcgccca 2520

ccaccacgcc cggctaattt ttgtattttt agtagagaca ggctttcacc acgttggcca 2580

ggctggtctc aaactcctga cctcaagtga tccgcccacc tctgcctccc aaagtgctga 2640

gattacagga gtgagccacc gtgcctggcc agtcctatcc tccccccaac ctttttttttt 2700

tttttgacat ggagtctcac gccatcaccc aggctggagt gcagtggcgc catcttggct 2760

cactgcaacc tccgcctccc atgttcaagt gatcctcctg cctcagcctc ctgagtagct 2820

gggactacag atgaagacaa gcacctgtgg tgcccctcac tgcaagaagt cagggaggca 2880

ttccacagcc tgggtgccca cagtcctgcc ctgtaccctc tggggccctt ttggcacggt 2940

ggcagcgctt ccagatttcc tttcagaaag atgcagtcct tccgcagctg ggccatgcag 3000

atctcctggc atccaacttg ttggaaacag tggccctctg ctcaatgatg agactgggga 3060

ggggaaacag gaggacattt caaggaaagg atggagcgtg gatatgaatg ggaagcgggt 3120

ggtgggggag ctcatcagca tcctcagagg ggctcatgcc cacgctgcaa cacagaatgg 3180

gacttgccag atgtttgtag tcgactctga gtgccccgtg ctgagaaacc tgaaagcaca 3240

cccacctatg gctgcgcgtg ttgcacgctc aaggctgagt tcacatagtt ctgtagcctc 3300

ctcctacacc aagtcaggtc ggccctgtgt gaccagtaga agagatggga tgtcactttc 3360

gagggtgctt ccaggcgagg ctggcctgaa tgagaatgag gagcaggacg ctccccaaga 3420

gattgccttg gacatcagcc tgggccacat ctacaagttc agacccattc agcagctaaa 3480

ttccaggagc atcacggaga atctccggcg agctcagcac caaggcaggc actgtgctgg 3540

ctgtgggccg gatgctacag aaactgtgtt atcggacacg ggtcctgatc taggccccaa 3600

gagagagttc ttgtacaaga aagaattggg gccaggcatg tttctggcgc tgtgtgccca 3660

ggcccagccg ggggcctaca ctgatgagaa cctcatggga ctgattgagc tgctgtgccg 3720

caccagcctg gacgtggggc tccgcctgct gcccaaagtt gacctccagc agcttctcct 3780

cttgctcctg gagaacatcc gggagtggcc agggaaggcg cttccttcca ggacagatgt 3840

cccacggctt gcagatggct gggcccagga gacggtgcta gcccttcctc tgagagaagg 3900

ggtgcaggct gccgccaccg tgcccatcct cctgtacaac ctggaggatg gcttgtcaga 3960

ccatcccctg gaccagggcc ccgctgccct gcccggcggc cctgcagccc tgcctcggct 4020

ccagctctca catcgccaaa gaagcccaag atacaggcac ctggggaaac cgttgacggt 4080

ttttgctgtt ccagttttgg ctgcttctca agatcacgaa gcccagggca ctctgcaagg 4140

gtttctgcaa gttcagcagt tcatcgctgg agggcgcttc cttccaggac agatgtccca 4200

cggcttgcag atggctgggc cccaggagac ggtgctagcc cttcctctga gagaaggggt 4260

gcaggctgcc gccaccgcgc ccatcctcct gtacaacctg gaggatggct tgtcagacca 4320

tcccctggac caggggcccc gctgccctgc ccggcggccc tgcagccctg cctcggctcc 4380

agctcccaca tcgccaaaga agcccaagat gcaggcacct ggggaaacgt ttcccactga 4440

ctggagcccc ccgcccgtgg aattcctcaa cccgagggtg ctgcaggcca gtcgggaggc 4500

cccggcccag aggtgggtgg gtgtggtggg cccccagggc ctgaggagac tggctggtga 4560

gctgcccgag gagttggagc aggaacacct ggacttggac ccgaagaggg gcctggcctt 4620

gccagagaag ctgttctgga acacgtcagg cctgagccag caggctgcgg ccccagagtt 4680

ttcctggggg ggctcaggaa gctacttcaa caacctggac tacttactgc aggagaagag 4740

ggaacaggcc ctggagcagg agcgagagag gctgcttctg caggagtgtc tcaatctcaa 4800

ctccttggat cttgatgaag aggaagtgcc actcacaccc gagcacagaa agaggcaaga 4860

gagctctctg gggccctttc ataagggtac caatcctatt catgaaggct ccaccctcat 4920

gcctcatcac ctcccaaagg ccccacttcc taataccttc accctggggc tttccttccg 4980

gagacaagca gtaaataaga tcagtgaagt tgtgctgcaa gggctcctga gaaaggctaa 5040

cgctgggggc ataaggagtg ctgggaaagg tgtgggctct gatgatgtgg gctctaatca 5100

tgtgggcttt gatgatgaag gctctgatga tgaagggatg ctggtggaaa agtactcagt 5160

gtccctgcag accatcccgc cggtccatcc aggtgagact gtgtttctgc ccaggtgtca 5220

ccccctgcca tgcatcctgg actcctcact cctgaagcca cgcagccacc tggaagggct 5280

gttcctcagg cagtatgctg agcattggga cctcaaggat gaggaagatg cagtctctgc 5340

cctagaggag cttacagcag caggaagttt ctgtcatagg acagacccag ggctcaccaa 5400

gactcaagca gatgatgaag cctggggctc actggcccaa tcagcgtatt cagactggct 5460

ggctgcttat gaggctcttg ggccagggct gcctgctcag tgggcagctg actctagctg 5520

ctgcaaaatg cctttcactc aaaggttttt gcttttgcca atcccctccc tcacatgcct 5580

tgaaacaagc actttcaaag acaaagacat aaacaacaaa agggtgcagg ctgagttgcc 5640

aacttacagt gtcattgggc cgattcaggt tcttgactgc tgcacaaaag aatttgagag 5700

caagtacaaa gcaaaagtag gtaaagaagt ttattgcaaa gcgaagatct cctgggaggc 5760

ccccgtggag aagaagactg agtgtatcca gaaagggaag aacaaccagg tgggtgcttg 5820

gacgctgctc ctggtgctgc cttcacccca ggacgtctcc tcccattctg gccctcgcgc 5880

tctcactaac cggacacctt tctgccccca gaccgagtgc ttcaacttca tccgcttcct 5940

gcagccctac aatgcctccc acctgtacgt ctgtggcacc tacgccttcc agcccaagtg 6000

cacctacgtc aacatgctca ccttcacttt ggagcatgga gagtttgaag atgggaaggg 6060

caagtgtccc tatgacccag ctaagggcca tgctggcctt cttgtggatg gtgagctgta 6120

ctcggccaca ctcaacaact tcctgggcac ggaacccatt atcctgcgta acatggggcc 6180

ccaccactcc atgaagacag agtacctggc cttttggctc aacgaacctc actttgtagg 6240

ctctgcctat gtacctgaga gtgtgggcag cttcacgggg gacgacgaca aggtctactt 6300

cttcttcagg gagcgggcag tggagtccga ctgctatgcc gagcaggtgg tggctcgtgt 6360

ggcccgtgtc tgcaagggcg atatgggggg cgcacggacc ctgcagagga agtggaccac 6420

gttcctgaag gcgcggctgg catgctctgc cccgaactgg cagctctact tcaaccagct 6480

gcaggcgatg cacaccctgc aggacacctc ctggcacaac accaccttct ttggggtttt 6540

tcaagcacag tggggtgaca tgtacctgtc ggccatctgt gagtaccagt tggaagagat 6600

ccagcgggtg tttgagggcc cctataagga gtaccatgag gaagcccaga agtgggaccg 6660

ctacactgac cctgtaccca gccctcggcc tggctcgtgc attaacaact ggcatcggcg 6720

ccacggctac accagctccc tggagctacc cgacaacatc ctcaacttcg tcaagaagca 6780

cccgctgatg gaggagcagg tggggcctcg gtggagccgc cccctgctcg tgaagaaggg 6840

caccaacttc acccacctgg tggccgaccg ggttacagga cttgatggag ccacctatac 6900

agtgctgttc attggcacag gagacggctg gctgctcaag gctgtgagcc tggggccctg 6960

ggttcacctg attgaggagc tgcagctgtt tgaccaggag cccatgagaa gcctggtgct 7020

atctcagagc aaggtaaagc tgctctttgc cggctcccgc tctcagctgg tgcagctgcc 7080

cgtggccgac tgcatgaagt atcgctcctg tgcagactgt gtcctcgccc gggaccccta 7140

ttgcgcctgg agcgtcaaca ccagccgctg tgtggccgtg ggtggccact ctggatctct 7200

actgatccag catgtgatga cctcggacac ttcaggcatc tgcaacctcc gtggcagtaa 7260

gaaagtcagg cccactccca aaaacatcac ggtggtggcg ggcacagacc tggtgctgcc 7320

ctgccacctc tcctccaact tggcccatgc ccgctggacc tttgggggcc gggacctgcc 7380

tgcggaacag cccgggtcct tcctctacga tgcccggctc caggccctgg ttgtgatggc 7440

tgcccagccc cgccatgccg gggcctacca ctgcttttca gaggagcagg gggcgcggct 7500

ggctgctgaa ggctaccttg tggctgtcgt ggcaggcccg tcggtgacct tggaggcccg 7560

ggcccccctg gaaaacctgg ggctggtgtg gctggcggtg gtggccctgg gggctgtgtg 7620

cctggtgctg ctgctgctgg tgctgtcatt gcgccggcgg ctgcgggaag agctggagaa 7680

aggggccaag gctactgaga ggaccttggt gtacccccctg gagctgccca aggagcccac 7740

cagtcccccc ttccggccct gtcctgaacc agatgagaaa ctttgggatc ctgtcggtta 7800

ctactattca gatggctccc ttaagatagt acctgggcat gcccggtgcc agcccggtgg 7860

ggggcccccct tcgccacctc caggcatccc aggccagcct ctgccttctc caactcggct 7920

tcacctgggg ggtgggcgga actcaaatgc caatggttac gtgcgcttac aactaggagg 7980

ggaggaccgg ggagggctcg ggcacccccct gcctgagctc gcggatgaac tgagacgcaa 8040

actgcagcaa cgccagccac tgcccgactc caaccccgag gagtcatcag tatga 8095

<210> 74
<211> 435
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (86)..(86)
<223> a, c, g or t

<400> 74
ggatgttcga tcactatagc cttgtgcctc tagatgctgc tcgagcggcg cagtgtgatg 60

gatgtcgcgg ccgaggtact tcccanggtg ggggccatat ttctctgtgt cctactctga 120

gcaacttctc agagatacga gggggctagg gttttcccat ctgggaaatg gggtgaaagt 180

ctgcagattg ttaaatgaaa tatagaatca gcagaaaaag aaaagtcagt gatataaata 240

gatcatttca tagaaattag ggtagatttt tatttcaact actactggag aatttaataa 300

aaggcattat ttgaaaagtt tttctaacat agatttaggg ttttttttt tagagtggac 360

acactacatt taaaagcaat tatttgtgct atccagattt tttattatct gaaaatgaaa 420

ttatctgttt tactt 435

<210> 75
<211> 608
<212> DNA
<213> Homo sapien

<400> 75
ggggggcata tttctctgtg tcctactctg agcaacttct cagagatacg agggggctag 60

ggttttccca tctgggaaat ggggtgaaag tctgcagatg ttaaatgaaa tatagaatca 120

gagaaaaaga aaagtcagtg atataaatag atcatttcat agaaattagg gtagattttt 180

atttcaacta ctactggaga atttaataaa aggcattatt tgaaaagttt ttctaacata 240

gatttagggt ttttttttt agaggggaca cactcccttc ctactcccca acaaggggct 300

cactatcccc aaagaaggag ctgtgggggga cccacgacgc agccccggta cgggattaca 360

gcatattctc atctcgggcc ccgaggctgc ctgtggggcg aggggagacc tcccatcacg 420

gagacagatc acagaccacg agtgcctttc ccggaccggg acgtggcctc cagagcaggc 480

accagctctt tccctctcta gacagaaata ttttggtaag gttctggggc agggagggag 540

catgaagtac gaggaaaact tgaattccag attcttaatg caaagtattt atcatttcta 600

ccagaaat 608

<210> 76
<211> 727
<212> DNA
<213> Homo sapien

<400> 76
gatgattcga ctcactatgg gcgaatgtgc ctctagatgc tgctcgagcg gcgcagtgtg 60

atgggtcgcg gcgaggtact tcctagacac tgtagtttaa aagaataatt tcctgtagaa 120

aaacctaatc taataggacc tactcataat ttggcaaaat catcctcata attccatgct 180

actgaaatgg gggtgtaata aagaaaaaat catttgttta ttcatttaac agggtttatt 240

aaagcaccaa tatatgccca gaactttgca gtgcactgga agttcagaga taagaaagcc 300

acagtttctg ccatcaaaat gctaggaatc tagaggtctc aaacatcaga ggtcccttca 360

aactatgtct accaggaaaa taaaatacct gatttcatcc ctgcaaacaa ctctatacag 420

acagaaatgc ttcctagctt ggaaactata gaaaattgaa catgaaaatg agagcaaaga 480

ttctacatca aaatggcaat gatccaatca gcccagtcaa ggctgagtgg gttgaatggg 540

gacttcgagt ttggattcaa tgctttgaac ttcatagctc cagagaagct gtgcaaaaag 600

gggggattct agggaatctg aggaagattg ttggagaaac cagctttctt cttgtctcct 660

aatccctgct aagattgata ccagtcaagg agctgagtta ctttcatttc attatgttga 720

ccaagag 727

<210> 77
<211> 3052
<212> DNA
<213> Homo sapien

<400> 77
ggaggtggag gggaggagtg cctgcatggc catgggattg ttcttcattc cctttctcaa 60

ctgcacccag cagcagtggt ttttgctagg ccttttgaag acagcaggaa tctgggagaa 120

ggaacatcat cgtctttcac agcatggaaa catcaatctt attccagaga agggaagaag 180

tccccaaagg tatgtccggt ttaacagttt ctcaagtggg ccaggaagtt ctttttcatg 240

ttctgggctc aatcgtgatg ctttgatttc acttggtatt ttacttttag ttttgtctct 300

aacatctgga gcaaagatca gaagacctga gttccagatt tattctgtga ctcaatcact 360

gcttcaatca ctgagggacg tggtatgatg ttctttgctc tgagcttctg cttcttgatc 420

taaaaagaca ggtggttgtt ctgggtgatc aaaatctctt cttgctctgt gtatcttaga 480

atccatgaaa ttatgtctct ttgcatctat gataacaaat caaaatatga atcacaattt 540

tcttacaggg ttgctgaggt taaagccgtg aaagcctggg aacaaatctc cccagggaag 600

gtttaagcaa gcattaactc tttcttctgc aagttccccc attcattgct agggtgttgt 660

ctgatagcca ccatgcacag ttgactttcc ttgggaaatc atgagtcttt tatttttgca 720

atttggcttt atctcctaca catcatccag atggtgcaac atctggtact atatgtccaa 780

tggaattgcc acaactcagc cactgagatc caacattcct gcctctcttc attccaagat 840

tccttctgtt ctttattgtg gtagacaaga gcaagcattt ttatttgcag agatgtactt 900

atttaaaaag caattgtgaa agcatttgat ttaataatta ggattgtttc tccacgattc 960

tttttctgaa aaaaaatttt tttggtgtaa atttagagtt cataagtaac accaaagact 1020

gagacaatga agtccatgca tactgtaaaa aaataacata aataacagtc cctagtatgc 1080

atcagtcatt gtactagagg ctttgtgtgt acattgattg tctcacttga aaatcaaaac 1140

acggcagaag agtcaaatgc ttttgaatgt gctgaaactc aaattttctt ccaggtcatc 1200

tctgtatgat tcctctttta cactgccatt actggcttct aaaaccacta aagctatgca 1260

gatagcctgg tcctaacatg ttctaaactg ggcattagtg aggggctgcc tgctggcttg 1320

tttaggggcc agttggctca gtttggaatg gtttaatatc agcaacaata gcattgggtt 1380

gatttgaata taaacaactt agactttaaa agttcatcct gaaaaacaag ccttcaagga 1440

caagtgagga catgaagcaa attcctaagg atgcctgggg ttcaggaagc aaagaagaat 1500

ctttggttat tcatgaaaac caaataccag ctatgtggca tcttctggga aagcacaggg 1560

gtgggagaat cctggggtgt gtttggcagc actgtccaaa gtacagttga cttcttaggc 1620

tgctgaacaa atttcttctc ttgccccagg agaatttgat ctgcaggttc ccataagtag 1680

agtaacatct ttctcttgaa ataggtgctg tgtcaaagtc tgtatcataa gcttctcttg 1740

gtcaacataa tgaaatgaaa gtaactcagc tccttgactg gtatcaatct tagcagggat 1800

taggagacaa gaagaaagct ggtttctcca acaatcttcc tcagattccc tagaatcccc 1860

ccttttgca cagcttctct ggagctatga agttcaaagc attgaatcca aactcgaagt 1920

ccccattcaa cccactcagc cttgactggg ctgattggat cattgccatt ttgatgtaga 1980

atctttgctc tcattttcat gttcaatttt ctatagtttc caagctagga agcatttctg 2040

tctgtataga gttgtttgca gggatgaaat caggtatttt attttcctgg tagacatagt 2100

ttgaagggac ctctgatgtt tgagacctct agattcctag cattttgatg gcagaaactg 2160

tggttttctt atctctgaac ttccagtgca ctgcaaagtt ctgggcatat attggtgctt 2220

aataaaccct gttaaatgaa taaacaaatg atttttttctt tattacaccc ccatttcagt 2280

agcatggaat tatgaggatg attttgccaa attatgagta ggtcctatta gattaggttt 2340

ttctacagga aattattctt tcaaactaca gtgtctagga agtaccttgc acaaaataga 2400

cacaggatgc ataattgttg attaattagt tcattttaac ctatacacct tggggttatc 2460

tagttatgtt ttcatgaaaa ctagggcttt gtaatatcaa tgtcatgtct aataagccta 2520

aatttccagt tctaaaaaca aagttattca acaacacttt ctcaatttga tgtgtttttc 2580

ttttcctgaa ataatttttt ttcttaagct tttgatattt gcttttgtgc tgacatttga 2640

tagtcaaaac tcaaacgatc tcaagtactt ctttctcagc tcaataggat catgtttcgc 2700

ttaacatttg tgaagttgat ttatctaatt cattttgatc ttgctaaaat atgaccttta 2760

aaatttttca tgtcacattg tgacccagaa agaccaaata catacatgag atcatagttt 2820

tacctgataa tatcactcct agtttgatga tggaattaat ttgtcaatat ttcatggagt 2880

tctgctgaaa cgattgcaat ttcttgggca agtttcagaa cgaccagcat atctgtggtc 2940

atgtcatctt gggtcatgtt aaccagttcc attggattat catgctgttc ccttgtctga 3000

caaaaggatc tttgtgaaag cacacactgt ttcttcccat aactcacatt ca 3052

<210> 78
<211> 416
<212> DNA
<213> Homo sapien

<400> 78
gccgcccggg caggtagcct ttttctctcc agccttgaat tgttccctgt tggcttccca 60

agggcccatc tgctggtaca gtccacactt ccaaagccaa gacccgagag ggctttcact 120

gccccaagcc tctctcctgt gaccttggga ttctgtcttg gcagaatcct ttgtcagcgg 180

ctcttgctct gtccttcctg tttggccaca gctctttcaa tcaatgggta ttctagaacc 240

gcaggatgtc agagctggaa gggacgcgat accggtttac acaaggggaa actcctcgag 300

gctctgggag ggacggaggg ttttggtgac agagcgagag ctaaaattga ggattcctga 360

atccagatct tgcctcccat cagccatctt tctcccaata aatttatgtt atgtgc 416

<210> 79
<211> 1451
<212> DNA
<213> Homo sapien

<400> 79
taaaaacttt gtcccgatcc atccagaaaa gagtaggtag ctgcatcctg acagcctggc 60

aaagtcaaga aagttgaagg agaaacatac ctttggagag ggggttttct ttaaaactag 120

tgttaagaaa tgcttaggga ttttttttte ttatttttca taactaaagc tttcacccag 180

agccggctct gtttgcactt tgctgccgac attgcaaact ttttggcagg gtgggagact 240

gagtctcatt ctgtcaccca ggctggagtg cagtggcccg atctcagctt actgcaacct 300

ctgccctcca gggttctggc aattccgcct cagtctcctg agtagctggg attacaggca 360

tgcgccacca cacccggtta atttttgtat ttttagtaga gaccaggttt catcatgttg 420

ggcaggatgg tcttgaaccc ctgacctcag gtgatctgcc catctcggcc tctcaaagtg 480

ctgggattac aagtgtgagc catcgcgccc ggcctgcaaa ctttttttgta ggtatttctg 540

gtaaacaaat ccttaggtta tctttgctgt ggttgtggtt tggctttagt catgatttca 600

aagtagaaat agctaggcat tatttttttga aatatatgac ctatatgtag tcaagaatcc 660

actgaacaga gggaagcaaa ccttttggaa actggctttt gggcagacag taaacgtcca 720

gtttgatgct ggaagcatga acagcttcat caggtaggta ctcctcaact ctgatgagtt 780

tgtcctttca gcctaagggg gtggaaggga gttgtttgag aatagcaaat acgcatgttg 840

attgcgagtg tgtggagaca aaggcagttc ccaccacagt taggtcctgg ccgttgtgtt 900

gtttcctcgc ctgcgatgct ccttgtacgt cctcaccctc ctctcccgcc tctgccttct 960

gctgggtcaa aggtggcctt tttctctcca gccttgaatt gttccctgtt ggcttcccaa 1020

gggcccatct gctggtacag tccacacttc caaagccaag acccgagagg gctttcactg 1080

ccccaagcct ctctcctgtg accttgggat tctgtcttgg cagaatcctt tgtcagcggc 1140

tcttgctctg tccttcctgt ttggccacag ctctttcaat caatgggtat tctagaaccg 1200

caggatgtca gagctggaag ggacgcgata ccggtttaca caaggggaaa ctcctcgagg 1260

ctctgggagg gacggagggt tttggtgaca gagcgagagc taaaattgag gattcctgaa 1320

tccagatctt gcctcccatc agccatcttt ctcccaataa atttttgttt tgtgcaaggc 1380

taaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1440

aaatgagcgg c 1451

<210> 80
<211> 1336
<212> DNA
<213> Homo sapien

<400> 80
ccgaggtaca aaggctttga ggtccatgga ctatacttgt ccccatttat catcccaggt 60

ggtgctttga ccctgccata ccctggctat taagataaaa agatttgtgg acattaaaat 120

tatgaatatg tcagtaataa tccagcacac attgaaatat tgacacagat taccataatt 180

tgtgcaacat cttataaaca atgtcatttc catagtagtc taaggcttca ccagcctggc 240

ccactgtatc tagactttag gttcatttta ataattatgc tttccttctc tgtatcattt 300

gggaagttga taaatatcac ttccttagat accttcattc agtgatatat ctggctttta 360

caattaaatt ggaaaaggta agtttctctt tggtgggttg agagttggac catcaattct 420

aatctacaaa aggaaattca tgatttcact ctgacgccta ggatctagcc aaggctggtc 480

tgcagtatca aatgtccaaa ctcatctact attagccata ttttgtgagt cgtttgtcta 540

aactttgtca aaaatgcctt tgccatgatt ttgttgctat ctggatttca aacatggaca 600

gttaggaaga tgtgcattga agtaggaaaa tttgttcaga ttgctgtatt tatttttac 660

attaaacatg gacatgtctc tcaaaaaaaa aaaaaaaaaa ggctgggggt actcggggcc 720

atagcggttc ccgggggggaa atgggtaccc gcccacaatc ccacacaaca accgagcacc 780

ccccccacc cccaaccaca atacatataa cacaaacctt acataaaaaa cactaacaca 840

acaaacacac aacaaaaaaa acacaccaac acacaaatca ccataacaca acaacataca 900

ccacaacaca caacaacccc gcccacccc caccactcaa caccaccaca caacacaaca 960

cccgcaccga cccccacaca cccccgcccc acccaccaca cgccccacca caccaccacc 1020

cacacacaca cacactccca ccacacacac acacccccc caacacccac ccacccccac 1080

aacaccaccc caccacacca cgcacccaca ccccacacac cgccaccacc acacccccac 1140

gaccacccac cacacacacc cacacacccc cacacccacc cacaccccca cccctcccca 1200

ccaccacacc acacccaaca tctcactcca ctctctcccc ccaccaccca cactcaacca 1260

catcatcact cccctctaca cacaacaaca tcaccaacac tccacctgca cacacactca 1320

ctccccacac atcacc 1336

<210> 81
<211> 1605
<212> DNA
<213> Homo sapien

<400> 81
atttataagg cccttcaaat ttgtggcttc ctttctcata cttctcaagt ataatgaaag 60

ggggagaaaa accccaccat caacacaaaa gaaggctata aagactgtgc acctttaac 120

aagtcaattt gtagtcagtc cctgggcctg tctttttttt tttttaattt tgaagctacc 180

tgaggtttag aattccttca gccctagctg cttttattct gctttttatt taaacaaaaa 240

gagggggagg atctgaagga aactagtttt ctgtacaaag gctttgaggt ccatggacta 300

tacttgtccc atttatcatc ccaggtggtg ctttgaccct gccataccct ggctattaag 360

ataaaaagat ttgtggacat taaaattatg aatatgtcag taataatcca gcacacattg 420

aaatattgac acagattacc ataatttgtg caacatctta taaacaatgt catttccata 480

gtagtctaag gcttcaccag cctggcccac tgtatctaga ctttaggttc attttaataa 540

ttatgctttc cttctctgta tcatttggga agttgataaa tatcacttcc ttagatacct 600

tcattcagtg atatatctgg cttttacaat taaattggaa aaggtaagtt tctctttggt 660

gggttgagag ttggaccatc aattctaatc tacaaaagga aattcatgat ttcactctga 720

cgcctaggat ctagccaagg ctggtctgca gtatcagatg tccaaactca tctactatta 780

gccatatttt gtgagtcgtt tgtctaaact ttgtcaaaaa tgcctttgcc atgattttgt 840

tgctatctgg atttcaaaca tggacagtta ggaagatgtg cattgaagta ggaaaatttt 900

gttcagattt gctgttattt atttttaaa ttaaaaatgg aaatgtaaaa aaaaaaaaaa 960

aaaaaaaaag gctgggggta ctcggggcca tagcggttcc cgggggggaaa tgggtacccg 1020

cccacaatcc cacacaacaa ccgagcaccc ccccccaccc ccaaccacaa tacatataac 1080

acaaacctta cataaaaaac actaacacaa caaacacaca acaaaaaaaa cacaccaaca 1140

cacaaatcac cataacacaa caacatacac cacaacacac aacaaccccg cccacccccc 1200

accactcaac accaccacac aacacaacac ccgcaccgac ccccacacac ccccgcccca 1260

cccaccacac gccccaccac accaccaccc acacacacac acactcccac cacacacaca 1320

cacccccccc aacacccacc cacccccaca acaccacccc accacaccac gcacccacac 1380

cccacacacc gccaccacca cacccccacg accacccacc acacacaccc acacaccccc 1440

acacccaccc acacccccac ccctccccac caccacacca cacccaacat ctcactccac 1500

tctctccccc caccacccac actcaaccac atcatcactc ccctctacac acaacaacat 1560

caccaacact ccacctgcac acacactcac tccccacaca tcacc 1605

<210> 82
<211> 952
<212> DNA
<213> Homo sapien

<400> 82
tatatatagg cgcatgggct aaatgatcat gctcgagcgg cgcagtgtga tggatcgtgg 60

tcgcggcgag gtaccaggtg aagtgattgg cctgcagtta gggctgtgtt gtgcaaaaat 120

cacttgtttt ggggtgttag aaccacattt aggcgagaag atcacttttg gggagcttgg 180

gaactgaggc aggccgcagg tgcagcagag gcatgagctt gcccgggccc gccctcttgt 240

ctatgctcat ggagtgaagg gggggccagc ggaatggccc aaacaactga tttgtttttc 300

ttttttaaa tctttttcag acaaatacca ttgtgtttaa gcgaaatgtg tgtataatgc 360

caaatcactg taccccacaa ccctgcacac ctctacactg gactgtaatt tcttgttcct 420

agtttgtctt gctagactgt aagctccgtg agagcaggga ccgtgtctgc ttgttgagtg 480

ggctttcccc tgtgcctgcc agcatgcctg gcatctagca ggtcttctgt aaagatggga 540

tgagtttgta aaccctccag cctcaggagt ggctccatcc tctgaggctc tggggccctc 600

tggcaggcta gtcatttttc tgccatgtac gtacaatgct ttattttcat gttgtatttt 660

cctttctagc cagtaagcaa agctcctact ggagcaagtt tctttgtggt gttatctctg 720

tatcctttca tcagggccag gccagagtaa gtaggtgctt cgaaggtgtt tggtaatgaa 780

tgactttggt accccgccca gtctggctca ctgtctgcgg ttgtctaatc tagatgaagc 840

ccgattcgaa ggggggattt ccggagtgct tggggttgga aaaatctctt attgccgggt 900

actcattggg tcgggcgggg gcgagtgggg actgctggtc ggggatgtcg cc 952

<210> 83
<211> 1933
<212> DNA
<213> Homo sapien

<400> 83
caccctttt gatagcatgt tatgaggctc accaaggtct atcctctggc agtatccctg 60

gcctgttctc cccatctccc ctgccctctg ctcaccagtt ccctaaatgt atcttgatct 120

cactagccct acagactgtg cacagagctg tttctacctc cagtgacatg cttccgccca 180

ggcttctccc cctcccgcct caatcttccc ctcaccaact taaggtctta ggggcccttc 240

taggaggcag tccccaactt ccagagccag ggtataggct acttctatgt gtcccatggt 300

actgccacag aagagtctca gctctctcca ggattcagtt ctaaggtcag tgcctaagat 360

aaaaatggag tgtaattaaa attcctctta gaaatctaag gaaggtgccc tattgaagac 420

caacatcttg aggtcccatg tagtcatttc ttgcccatgt gggaacacat tactgtttgg 480

ttgagtacca ggtgaagtga ttggcctgca gttagggctg tgttgtgcaa aaatcacttg 540

ttttggggtg ttagaaccac atttaggcga gaagatcact tttggggagc ttgggaactg 600

aggcaggccg caggtgcagc agaggcatga gcttgcccgg gcccgccctc ttgtctatgc 660

tcatggagtg aaggaggggc cagcggaatg gcccaaacaa ctgatttgtt tttctttttt 720

taaatctttt tcagacaaat accattgtgt ttaagcgaaa tgtgtgtata atgccaaatc 780

actgtacccc acaaccctgc acacctctac actggactgt aatttcttgt tcctagtttg 840

tcttgctaga ctgtaagctc cgtgagagca gggaccgtgt ctgcttgttg agtgggcttt 900

cccctgtgcc tgccagcatg cctggcatct agcaggtctt ctgtaaagat gggatgagtt 960

tgtaaaccct ccagcctcag gagtggctcc atcctctgag gctctggggc cctctggcag 1020

gctagtcatt tttctgccat gtacgtacaa tgctttattt tcatgttgta ttttcctttc 1080

tagccagtaa gcaagctcct aatgagcaag tttctttgtg gtgttatctc tgtatccttt 1140

catcaggcca gcccagagta agtagtgctt cgaaagtgtt tggtaaatga atgaattatg 1200

gtaacccagc ccagttctgg catcactgtc tgcaggttag tctaattcca gatgaaagcc 1260

ccagattcga aaaagggaag attattcaga agtgcatatg gttgtataag aaaaaaaaat 1320

ctcattaaat agacctgagt aaacttcaaa ttttgatgat ctggccatgt cggagctgaa 1380

gcttaccagc tctctaaagc atttgctagg gaatcatgca gatggtccaa aatatttttt 1440

agaaaagcat tcatctgtcc cgggatagaa gtagaggaag ctggtaggca ggcgctagac 1500

cctaaagaag gactgctgag ccagttctta agccagtggc tcttttgcct aaagcagcat 1560

tcctcaaact tctcactcga agcattaagt tacggtgaga tgctaacaca cgctttagac 1620

ttaagtcaaa atttactgtg aagccacagc aaaccatggc ttcacagatg aattagttgc 1680

tacacaaata agttcgatac ataagtacag taattgcttt taaaccttat aataagcagt 1740

aaattagaag ttaaaccatg tttttctttt acgaactgaa agaagagaat gcttttgata 1800

ctgagaatcg cacactgtgt ccaggcagat tgttcttgca taaccagacc cacaataaat 1860

atctacaact tttcttacat gtttataaaa taaatcttag aaaaaaaaaa aaaaaaaaaa 1920

aaaaagggcg gcc 1933

<210> 84
<211> 376
<212> DNA
<213> Homo sapien

<400> 84
gcgtggtcgc agccgaggta ctccactcca aatttcccaa gaaattcaga agaattgtga 60

acaagttgct ggtttcacaa tactgcaagg acactgcaag gttattccaa gttcctcagc 120

aagtgtttac acattgggcc aaggacagat ttttcctgga gaaggatttt accactgcca 180

ccatcttgaa attcttcatc gttttggaac acagagccat agattttcat ttctgcactc 240

agctctgttc tgagaccggg gccatagggg ttcttggaga gacagggcag atggaagaag 300

tggaaggcat ctgccacact gtagagtgcc tttaagccac ccccattgcc tgagttgttt 360

tccttttta caaatg 376

<210> 85
<211> 1325
<212> DNA

<213> Homo sapien

<400> 85
ggccgcactt ttttttttt ttttttttc agagacggag agatttattt taaggaattg 60

tgtggctggc aagtccgaaa tccatagggc aggctggcag gctgaaaagt taagtgaggg 120

ttaatgccat agaccggagt cctgagtttg aattccacag ggcaacaggc tggaaactca 180

gatagggttt ctatgttgta gtctcaagga gaattccatc ttctttagga aacctcaatc 240

tttgttctta atgccttcaa ctgattggac agggcccacc cgtagtatgg tgggtaacct 300

gctttactga tttaagcgtt aatctcatct aaaaaaatac ctttaccacc atgtccagac 360

tagtgtttga ccaaacaatg agttggcata ctccacgcaa ggtgacgtgc aaaatgaacc 420

atcacagcac ttcttgtgtc ttcatgtgcc cacctggtac atgggggtac aactatctgg 480

cccaaagggg tggtcaatcc tgagtggata aataatagga atttagtcct atgatgtgtc 540

ttgatttttc ttttcttttc ttaccttcct atttttatta acccgggtct tctgcgccca 600

aagactcagc ctcattcagg accatattat gttcatactt ctgctgcgtc caaggagtgt 660

tgacgaaaaa gtggggcctt ggaggggtag aggccaaggg acagtttccc ctctgccctt 720

tgaagttcac gatcttccat gcaacaaaat tgttttctgt gaaaagcagg aaatgaataa 780

caacagcgta ggtgcgttgg ctatgtccgg tggcatttct tcagaatttt cattaatgac 840

acctgatttt ggaggcattg tatattttta aatacatcca gatgttgttt cagttgcttc 900

ctctttggtt ctttttgctt ttcgttgttg agcgtcactt aaattcgtgt catttcatgt 960

tggtacaggt actccacttc aaatttccca agaaattcag aagaattgtg aacaagttgc 1020

tggtttcaca atactgcaag acactgcaag ttattccaag ttcctcagca agtgtttaca 1080

cattgggcca aggacagatt tttcctggag aaggatttta ccactgccac catcttgaaa 1140

ttcttcatcg ttttggaaca cagagccata gattttcatt tctgcactca gctctgttct 1200

gagaccgggg ccataggggt tcttggagag acagggcaga tggaagaagt ggaaggcatc 1260

tgccacactg tagagtgcct ttaagccacc cccattgcct gagttgtttt ccttttttac 1320

aaatg 1325

<210> 86
<211> 744
<212> DNA
<213> Homo sapien

<400> 86
gcgtggtcgc ggcgaggtac ttttaactta aaaaaatgaa catctttgta gagaattttc 60

tggggaacat ggtgttcaat gaacaagcac aagcattgga aatgctaaaa ttcagttttg 120

cctcaagatt ggaagtttat tttctgactc attcatgaaa gtcattcctt aattggaggc 180

ccaccattca attattcatc tattaattcc ttgatccttc atttatccat tctgcaaact 240

tttcttgagc accagcacgg gtggccattt gtggacttct cttcattcct atgtgttttc 300

ttatcaaagt gatccactct cgaaaggctc ctttccagtc tgtggttggg ttcaagtcat 360

gccagggcca gggggcccat ctcctcgttt agctctaggc aaaatccagg ggatctgcag 420

tggggagcgg gggcaggaag ctggagggaa ggcctgtgaa gggtagggat gtggaaagac 480

aaggtgacag aaggacccaa taggaccttt ctatatctct ggcttagcat tttctacatc 540

atattgtaat cgtcttattt gctagttttc ttccttactg tgagtgacta acagtcatct 600

ttatcccaag tgcctggtac ataataagtg atcaataaat gttgattgac taaatgtaaa 660

aaaaaaaaaa aaaaaagggc tgggggaatc agggccaagg cttgtccggg gtgaaattgt 720

ttccccccac aaaaaaaaaa aaac 744

<210> 87
<211> 1833
<212> DNA
<213> Homo sapien

<400> 87
tgctccagga aagttctgtt actccaggga ctctctcttt tcctgataac atggccagca 60

agaaagtaat tacagtgttt ggagcaacag gggaagctgg tggcagactc cgccaagcac 120

ctgggtctga agcacgtggt gtacagcggc ctggagaacg tcaagcgact gacggatggc 180

aagctggagg tgccgcactt tgacagcaag ggcgaggtgg aggagtactt ctggtccatt 240

ggcatcccca tgaccagtgt ccgcgtggcg gcctactttg aaaactttct cgcggcgtgg 300

cggcccgtga aagcctctga tggagattac tacaccttgg ctgtaccgat gggagatgta 360

ccaatggatg gtatctctgt tgctgatatt ggagcagccg tctctagcat ttttaattct 420

ccagaggaat ttttaggcaa ggccgtgggg ctcagtgcag aagcactaac aatacagcaa 480

tatgctgatg ttttgtccaa ggctttgggg aaagaagtcc gagatgcaaa gattaccccg 540

gaagctttcg agaagctggg attccctgca gcaaaggaaa tagccaatat gtgtcgtttc 600

tatgaaatga agccagaccg agatgtcaat ctcacccacc aactaaatcc caaagtcaaa 660

agcttcagcc agtttatctc agagaaccag ggagccttca agggcatgta gaaaatcagc 720

tgttcagata ggcctctgca ccacacagcc tctttcctct ctgatccttt tcctctttac 780

ggcacaacat tcatgttgac agaacatgct ggaatgcaat tgtttgcaac accgaaggat 840

ttcctgcggt cgcctcttca gtaggaagca ctgcattggt gataggacac ggtaatttga 900

ttcacattta acttgctagt tagtgataag ggtggtacaa ctgtttggta aaatgagaag 960

cctcggaact tggagcttct ctcctaccac taatgggagg gcagattata ctgggatttc 1020

tcctgggtga gtaatttcaa gccctaatgc tgaaattccc ctaggcagct ccagttttct 1080

caactgcatt gcaaaattcc cagtgaactt ttaagtactt ttaacttaaa aaaatgaaca 1140

tctttgtaga gaattttctg gggaacatgg tgttcaatga acaagcacaa gcattggaaa 1200

tgctaaaatt cagttttgcc tcaagattgg aagtttattt tctgactcat tcatgaagtc 1260

atctattgag ccaccattca attattcatc tattaattcc ttgatccttc atttatccat 1320

tctgcaaact tttcttgagc accagcacgg gtggccattt gtggacttct cttcattcct 1380

atgtgttttc ttatcaaagt gatccactct cgaaaggctc ctttccagtc tgtggttggg 1440

ttcaagtcat gccagggcca gggggcccat ctcctcgttt agctctaggc aaaatccagg 1500

ggatctgcag tggggagcgg gggcaggaag ctggagggaa ggcctgtgaa gggtagggat 1560

gtggaaagac aaggtgacag aaggacccaa taggaccttt ctatatctct ggcttagcat 1620

tttctacatc atattgtaat cgtcttattt gctagttttc ttccttactg tgagtgacta 1680

acagtcatct ttatcccagt gcctggtaca taataagtga tcaataaatg ttgattgact 1740

aaatgtaaaa aaaaaaaaaa aaaaagggct gggggaatca gggccaaggc ttgtccgggg 1800

tgaaattgtt tccccccaca aaaaaaaaaa aac 1833

<210> 88
<211> 251
<212> DNA
<213> Homo sapien

<400> 88
ttaaaaatgt aaaaaaatga aaaaaaaagt tttgagcatt atttgcatca ttgggataca 60

tatgtcactt cacaagatgt tcaatttgaa ggaaatacca ctcattctct atgtcctgtt 120

gtctgtagtg tgcttcagtt ttctcatatt gagttgacct aaatcctgga ttcatgacaa 180

gaaaggagta agtactacta ttcattgttc tatttgttta taatctgtat tataaaattg 240

cacataatta a 251

<210> 89
<211> 458
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (327)..(327)

<223> a, c, g or t

<220>
<221> misc_feature
<222> (435)..(457)
<223> a, c, g or t

<400> 89
ttgatattta tttagaaaac aagggaaagc ttttaattgt gtgcaatttt gtaagacaga 60

ttataaacaa atagaacaat gaatagtggg acttactcct ttcttgtcat gaatccagga 120

tttaggtcaa ctcaatatga aaaactgaag cacactacag acaacaggac atagagaatg 180

agtggtattt ccttcaaatt gaacatcttg tgaagtgaca tatgtatccc aatgatgcaa 240

ataatgctcc actctgtcgc tcaggctgga gtgcagtggc gcaattagaa ctcactgcag 300

ccctaaccct cttggcttaa acaatcnctc caccttagcc ctctgagtag ctagcactac 360

aggcacacgc caccacaccc agctaacttt tttgaatttt tggagacagg gttcacctta 420

ttttttctcg tgccnnnnnn nnnnnnnnnn nnnnnnnc 458

<210> 90
<211> 251
<212> DNA
<213> Homo sapien

<400> 90
tggtcgcggc gaggtacaaa aacatggcag taggacaact ttgagctcaa cgcgcgcacc 60

tccccaggac ctctcacacc cacctgaccc atccaagggc cacaccaccc cgacagatac 120

tcccaccacc tttagaaaag agtcacccaa tctggagaaa ggtgtggagg ttacatcttt 180

agaaagaaat cattttaaat acatgaacat tagagaacac agtaaccgtg cttccaccca 240

gcacggggag g 251

<210> 91
<211> 2399
<212> DNA
<213> Homo sapien

<400> 91
ggccgcatct ttttattttt ttttttttte ttacaaatgt gagttttatt tgcattctaa 60

cccaagcaca gatcccacat acagacttct cgtccgcact atggtacagt aacaagtgca 120

aaatatgctt tatttcaggt acaaaaacat ggcagtagga caactttgag ctcaacgccc 180

acctccccag gacctctcac acccacctga cccatccaag ggccacacca ccccgacaga 240

tactcccacc acctttagaa aagagtcacc caatctggag aaaggtgtgg aggttacatc 300

tttagaaaga aatcatttta aatacatgaa cattagagaa cacagtaacc gtgcttccac 360

ccagcacggg gaggctgcag agaggcaccc caaacagtcc ccttctcttc tgtggttaac 420

caagcaagcc ccgcaaagct tttcccagca gagcacaccc agccagcaat gaggcctcag 480

gacagagcca gcactgtgac aaaagaactt gtcccaccta tctgagagca tgagatctgc 540

aagtgggatc cggactggac accaggaagc tcaacctggc tcgggccatc agtagctgtg 600

ggaccctggc ccccatctaa cctcccagtt gacttcctga aaactgggct agtgggggcc 660

tccccattgc tttagggagg atcgagtgag aagctgacca aatgcgcaga acacaaagcc 720

tagcctgtta gccccagctg ctatcactac tgttactcac gaagaaagga actgcagtgg 780

ctgctgtctg cccctggagt gaagacacca ggggtgcaca gaggctggac aagccacagg 840

gggctggatg gccacagccg gacctcggac ccatctcccg tgggactcct tccatttgag 900

gctgtccatc ctccaatggc agatgcagaa ttctcaggcc caggataccc ttatggaggg 960

ccgggggggcg gcccctcaaa accaccaagc tgaggccact gtgagcagtc ccctttgtag 1020

caggaagagc agctggacac atgggaagag gcttgcaggc tccagagaag ccaaaaggtc 1080

ctcagtcccc aaggtactag gtggaggtga gatcaggctg cccctcccag gacaagagtg 1140

tgagctggcc aggaggccac tgctgcccac cactgctcaa agtccttctc cacgcatcaa 1200

ccctgaggga cctggccagg ggatgcagac caaaggccca gcgggtcccc aggaacagga 1260

agcaggcggc aagaaggaaa gcaaagggct ctgcattccc accacgaggc ggggtggact 1320

ctgcaggagc cagctgaagc accagaacct tccaaggggg gctcgccagc cacacaggac 1380

acgagatgcg tgataggcag ggttaatgca gaaaccgctc atctgaatgc ttcccctgct 1440

tccaagcgca aaccaagtca ttttattctt ttaaaaaagc ccttactgtc tgaggctttc 1500

tttaaataat ccaacaggga tggaggtaat ggaggtctgg ctcagtagtc agcagacttt 1560

ttctgtaacg ggctagacgg taaatatttt gggctttcag agccaagagg caaaatcaat 1620

attatgttgg cgtaagagca aacaaatttc cacaattttt aaattgataa aatccaaaat 1680

acaataattg agtacatatt tttggtaaat acatgtctac taaaaataag aatttttttt 1740

tttgtgatgg gggagaaaac acttcattgg gttcaaagat aatattccct atcataaatc 1800

agtggcataa acagccagtg ggccaggtgg gcccaccgta tgctagctga caagggcagt 1860

cacacaccgg taagtgctga tacaggatga ggggtgcgga gctgaggatg ttgttctctt 1920

tgatgaaggc ctgaaaactt cccctggccc ctgctgccca cctcagggct cctcaggacg 1980

aagggcatga ggacaggttc ccactctggc ccttcgctgg gggacatggt ttggtttccc 2040

catcgccagg ctggccagct ggggcccaag cactgaccgt ccttccccac cttatcccat 2100

cagctgccct gcacaggagc agaaaaccca tgcctcagag ctacttccta aggacactgc 2160

ctctagaagg ccctaaagtg gcagagccgg ggatggcagg tgcagccctg gccactagca 2220

cagcagttgc cctcagatct gttagctgtg acccatagat tgggggaggg aggaaagcca 2280

ggaggagctt ttgcagaaag gtctggaagc taagtggggg tttcaggagc tctcagggtg 2340

cccctgggca ggtcccaagg aggcctgggg tgcctgacca gcacctcact ggccctcaa 2399

<210> 92
<211> 595
<212> DNA
<213> Homo sapien

<400> 92
gtgatcgact cactatagca tgttatctag atgcatgctt cgagcggcgc ctagtgatgg 60

ataggtcgcg gcgaggtaca aatcattaag gcaaatctac aaaggaccag taaggcactg 120

gggcacagca gctgactgca ggaagtctat gtaacttccg gtagtaaaac agcctttaat 180

gcagacatag gtgtgaaaaa ttcagtgtta tgtttttttt tcagtgtcaa cctgcaccta 240

aagcacaaaa ggtcacttca agtcacttca atgtaataaa caattcccat tcatcacatt 300

accagtgaaa agagaaataa ttcactgggc tagaaggaat tagaagcagg aataaagaaa 360

atgccctttg gtttacactt actaaagaga atgtctttaa ctctccaagg aataccattc 420

cgctttgtga aattatacca attaattagg actcctctat attcccatta atgtcggatc 480

attacataac ttcaatagaa taaacacttt tttcccactt tgctataata atagctaggg 540

atacctcaac aatataataa tggtttaaat tatgaccatt tctctttggc cttgt 595

<210> 93
<211> 1457
<212> DNA
<213> Homo sapien

<400> 93
cttatcctag agaataactc tgtatgaata aaattgctta attgagtctc ttactaaata 60

agtaactagt gccatgcttt tgtgagctct tggtatggcc catattactt tgtttttttt 120

ttttttattg ttgttttgtg atagtcttgc tctgtcgccc aggctgcagt gcagtggcac 180

aatctcagct cactgcaacc tctgcctcct gggttcaagc aattctcctg tctcagcctc 240

ctgggtagct gggactacag gtgcatgcca ccatgcctgg ctaacttttg tatttttagt 300

agagacaggg tttcaccacg ttggtcaggc tggtctcgaa ttcctaacct caggtgatcc 360

acctgccttg gcctcccaaa gtgctgagat tacaggcgtg agccaccgcg cctggcctgt 420

ttgtttttt aacatgattt ttctctaagc ttaaatacca caaggccaaa gagaaatggt 480

cataatttaa accattatta tattgttgag gtatccctag ctattattat agcaaagtgg 540

gaaaaaagtg tttattctat tgaagttatg taatgatccg acattaatgg gaatatagag 600

gagtcctaat taattggtat aatttcacaa agcggaatgg tattccttgg agagttaaag 660

acattctctt tagtaagtgt aaaccaaagg gcattttctt tattcctgct tctaattcct 720

tctagcccag tgaattattt ctcttttcac tggtaatgtg atgaatggga attgtttatt 780

acattgaagt gacttgaagt gacctttgt gctttaggtg caggttgaca ctgaaaaaaa 840

aacaaaacac tgaatttttc acacctatgt ctgcattaaa ggctgtttta ctaccggaag 900

ttacatagac ttcctgcagt cagctgctgt gccccagtgc cttactggtc ctttgtagat 960

ttgccttaat gatttgtaca aatgactggg aggcggggat gctgcctgtg tcctggtgaa 1020

ccttaatgaa ggggccgtct taggcacagt gcaaaacaag catttgtcct gtactgttag 1080

agccaaaatt gtgatgagca atactgataa ttgtccagtt tatgtcatct ttcccagatt 1140

ttaaaatctg ttctagatat tcttagcttg aaccactttt gattgtgaaa tgtattaggt 1200

gttgtcccat tattactgta aaatgaagtt ttgaatcttc ttgttaataa actgtggatt 1260

tcccctctca atttcttaaa caacaacaaa aaaatgcttg aagattgtct ttgagtgtaa 1320

gatctgcctt ttcagaaagg gagtgttagt ttgtaatgtt aaaaaataaa gacctcattc 1380

aataaaagtt gaagtcatct tttaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1440

aaaaaaaaaa actcggc 1457

<210> 94
<211> 936
<212> DNA
<213> Homo sapien

<400> 94
tatcactata gcgatgtgct ctagatcatg ctcgagcggc gcagtgtgat ggattgctgg 60

ctgccggcga ggtacacaac gccttctaat ttgatcttca tgactgtaga cgggtgcccc 120

attccccctg aaacaacggt tgcgaggttt gcgcattact gcttcgcctg tgcatgacta 180

tcttgctgta actactagcg atgatcttgt atcgaattgt atctcagcca cttatcctct 240

cagtatagta caattaccat tgcatacttg tatacctcat ggagaaggga gttctctcca 300

gtcccaaccc atgaatgtgt tatcgattcg cgcattcgtt cgctcgcaac ctcagataaa 360

cactcttcat actttatcac taccttccat cttagtcaaa gctcaagaca aaaattaaac 420

aaaatgacac ctagccattc acctattaca accatcgacg tataagacac accattcgac 480

gtgaggaaat gacacacgta tttccatcaa cgagagtgct ggatgggagg gcaggacaca 540

gactggaagt cagcttcagc gtatacccac ttactcacaa gggctgacca tgatggccct 600

gctcttgcat gaggactcac atggacagcc agtggcactg atgcctgcag gactgcatgt 660

gccgacaaga tgatgcacat tttaatattc aaaatgtaaa caaagtatcc acaaatagtc 720

caaaaaagcc aaagaatggc tttacttaac aaatttttaa attgccctta aaagtgactt 780

tttggcataa tgcctttctt aaaaaaagaa ctctcacaaa aaaccatttt ttttaaccca 840

aaaacaatgc ctagatcact tcagaactga aatcctcaga tgaaacaggt gaacacttct 900

gtgactgcta cctctcagaa ggcttgggcg tatcat 936

<210> 95
<211> 480
<212> DNA
<213> Homo sapien

<400> 95
tctcgagccg cgccagagtg atggatgcgg ccgccgggca ggtacttatt aagcagagca 60

ctttgtaaga ttcagaactg actcctgatt atcttaagcc attaaaagac tcaagttttg 120

catggacttt ttgtcctctt tggcccctga gtgtgcccca tctctgccca gcactaataa 180

cacgttggaa gagcaaagga tttcccacca gcgtgccaga tgatttatat aggcagggac 240

cttcattgaa aagaagggaa gggacactgt cttaaatgca catttaaggt tctttctgat 300

tattggcaat ctctatgaac caacacttaa tccatgggct aacagagaga ttttttttt 360

tttaatgtga agaggattaa agaataaaga aaaaacaaaa aagtctttat tctaaaataa 420

gaaatcagcc ccatctgtgg cacagttctc atgcagaata ttgcacccag tgtgaactag 480

<210> 96
<211> 1111
<212> DNA
<213> Homo sapien

<400> 96
cgtccaatca gaggctctgt atttataaat aacccgtgtg ggattatgct ctccagtgaa 60

tataactgga gcctaaattc acaaccttaa gatctgacag ccagatgtgg aaggtctaca 120

aagagctctg ctaataagta atatgtttgc aaagtgctct gctaagtaag gtacttatta 180

agcagagcac tttgtaagat tcagaactga ctcctgatta tcttaagcca ttaaaagact 240

caagttttgc atggactttt tgtcctcttt ggcccctgag tgtgccccat ctctgcccag 300

cactaataac acgttggaag agcaaaggat ttcccaccag cgtgccagat gatttatata 360

ggcagggacc ttcattgaaa agaagggaag ggacactgtc ttaaatgcac atttaaggtt 420

ctttctgatt attggcaatc tctatgaacc aacacttaat ccatgggcta acagagagat 480

ttttttttt taatgtgaag aggattaaag aataaagaaa aaacaaaaaa gtcttatact 540

aaaataagaa atcagcccca tcttggcaca gttctcatgc agaatattgc acccagtgtg 600

aactaacgct agaagcttca aactgtataa atttaaatgt atttgcatat tataaaaata 660

aagataaaca tatacatatt ttacactagt tatggaacag caatgaacgt cagtcgatcc 720

ctctttcaca tttaacagaa ctgaaatctg agtgctctaa atactgccac ctgtactgta 780

actatggctt atatgtgcac ggaaaacaaa atccctgaga agccattcga ctttttttt 840

tttttttct tttcttcaag tagcgcgctc cttggaggat cacagttctg aggttcaggt 900

tgtaaaacat ttgctccatg ttctcgtcca tgcttccccc caccaccccc tccccacctc 960

ttccccagtc atccaaaaag caccctgcaa gcacgcgttg tcactcaagt tcacagaaca 1020

cgctggggtg agtgcagagg gtctgccagg tgcaaaagat ggtccaggtg ttcagatgct 1080

ctcttttctc catggaaatt ccacagccac a 1111

<210> 97
<211> 395
<212> DNA
<213> Homo sapien

<400> 97
gcgtggtcgc ggcgaggtac caccagctac ttcagagtga agaaagaaag tgtctttcgc 60

cactggttct taaggagtgg atctccggtc cacatctccc aatgctgctt agcccaggaa 120

atgaagacag ctgacaacac agaggaacat taacgaaaag aacaaacccc aaacttgaaa 180

tgcataactg ggactgctgg aatgggccaa ggcataccac ggcgggtcac tgtcaccagg 240

agggcgcctg tgttttagaa gggagtggcc aacaccgtct ggccaacctg gaagggagcc 300

aacgtgacag ctaagaacac tagacaagca tcctagcctt ctcacccatc gatctcagtg 360

tgatcttgcc acagagccac tggcaagcca ttcgg 395

<210> 98
<211> 3813
<212> DNA
<213> Homo sapien

<400> 98
tttttttttt gcatacagaa ttattttatt taacttaaac catgtagtac tttactagaa 60

aaaagcagag taagagaaac taacgttgcc ttagcttcag ccattcaaaa tagacagttt 120

ctttttttcca ttatgtaaag aatccagagt atatcgcaat aacaggaata aattcttaca 180

acagaatata caaaaacatt ttgaaatttt tttcatctac tgatttttta tataaacagg 240

atttttagg aataatttat acacagaaag tcattttatg taacaaattg gccatgttat 300

tacctttttt tttcttactt aaaaaaaatt ttttttaac aagaaaactc agaaaatgca 360

ttatttgcgg tgcatccatt ccatcccgcc ttctggtttg atttttttta tcccagacaa 420

agggataccc agaggtagac aaactctggc aaacctctca cctcaacctc actggcttag 480

aaagcagaca ggtgttttca ccgggcgtct cttccaccgg tggatgtgtg tgtgcaacgc 540

caaacatcca aatgaaaagt tgaaaacaaa acccaaatag tttccagatc tttctcctat 600

gtagcgcaaa aaccccaggt gtttcccatg atagagatat tgtggagttg gggagacatg 660

gccaacaatc cactggagtt agaggaagtg gagtcaactt taaaaccaat ttttgtgtat 720

gttatagttt ttattaaaaa ttcttttctg tattggaaat acgtatcttc ggggttctgc 780

cttaaataaa atagggtctc ttctgtgttc taacaaagca gtaatcaatc atgatacgag 840

gcatgtaagc aaatggcctg tgtgccccat agctgtcatc gtttatgtga tcactgtgta 900

tgctgtaatc atcatgaaaa ccattggctg tgcagtcaag tcttcccatg tccatggcac 960

tgtgctgaca atgacggcag ccagaccaat caagccagat gccgaacggt gccatgcact 1020

gctgggctcc aacacctcgg ctgtttccat ggccaatgag gttctttcca tctgtaaagt 1080

agctacactc caaagggaaa tgaccagtga aaaatgatcc ccgctcccaa ccagagggtt 1140

ctgggaagac ccagatactg aagatacaga actacttgtt aaatcatcat cattcctttg 1200

ttataaacta gggccaaaaa gcatgggggg aaaagagggt gaaccacgga tggggtcagg 1260

gagaaatcct tttttggcat attgtaaagc tagctagtaa acaggtgtcc taaaaaaaaa 1320

aaaaaaatgg atcacaaaca cacatttgca cacacagaga aaaaaaaagt gtctttaaat 1380

tattttcggc aattataaac tacaaacatt ttataaaatt attctgttct tacaaaaatg 1440

caatgaaaca tttaattagt tcttgtaaac cagatcttcg tcaactgact acccgtaatc 1500

tactggactt aagagcccta ttgaaaacgc taatgagtga ctagttaaaa caatgttgac 1560

cagaaagatg cggacacaca cacgtgggac ataaaaatga gtgatcgtta tgttctcgca 1620

gaaacaaaac gctaaagtca tgccccagaa ctgacaactg acgagtgcga tgtagtttgt 1680

gttcttgaac aattcgcctt aactccgaga tcagtttcca gagtaggatg ttaggtttct 1740

ttgttgaaac caaatgccca cggatgactg ctgagcctcg atgtggtctg atagcttcaa 1800

gaacaatcaa gttcaaaacc ttgcccacat tccttccaag ccctgcaccc tgctgctcgt 1860

aactcttctc tgtgccagga tccctgtgac aatggttctg tacccaccac ccggtcattc 1920

caccatcaag gagcactcac ttctccaaat gacgcaacca gtcatctaag atgaaggcca 1980

aaaatgaaac cctctgacct gtccaaggtc tgtatcttca gcatctctgg caacagaggt 2040

tcagtctaaa taacatcctc tcttgaagtt ctcatttgta ccaccagcta cttcagagtg 2100

aagaaagaaa gtgtctttcg ccactggttc ttaaggagtg gatctctggt ccacatctcc 2160

caatgctgct tagcccagga aatgaagaca gctgacaaca cagaggaaca ttaacgaaaa 2220

gaacaaaccc caaacttgaa atgcataact gggactgctg gaatgggcca aggcatacca 2280

cggcaggtca ctgtcaccag gagggcgcct gtgttttaga agggagtggc caacaccgtc 2340

tggccaccct ggaaggagcc aacgtgacag ctaagaacac tagacaagca tcctagcctt 2400

ctcacccatc gatctcagtg tgatcttgcc acagagccac tggcaagcca ttcggctgcc 2460

agcctgcaga ggtggactgc tgtgggaaaa gcttcagtgg aacaaactgg cctagaatct 2520

ggtgccagag aatagttggc tcataaatct tttatcccaa agcaccccaa tgtgcaccga 2580

gcatctcgcc atttggtagt tccacagtga ctgctcttct attttacgaa gccacttttc 2640

gtaccattga aatgtgactc cgcatcttaa gaaacaggcc aggagaggcc aggtcggcca 2700

cgaacacggt cttccatctt ccaaggagga ggctcgttcc aaaggtatgt caaacgtcat 2760

gacatcccac gaaggagacc tggggtggtc tgggtcttga tcaagttcaa cagtgtctcc 2820

agtacacacc agctcctgct gcctcaggga agggctgtgt gggagaacgt cttaatgctc 2880

acatcaattc catgatgtgc ctcagtggtt catatgactt caagcagtgt gagctgctgg 2940

cagaaggagg ttgcatggaa aatgccacaa ctcagcttaa gattatttgg aacacacaga 3000

gaaaaatcaa gaggcagaga aatcagcatg aagttaggct agaactgtgg atagatgatg 3060

ccttggtata gtcaaggtgg gtgggttggt ccagtaaaat gactgcacca tcacacaagc 3120

caagttttct cttttattga ggtgtttctg aagcatttgc tatgatgggc actacactgt 3180

tagtgttttc tacttaaaag ctacctacat aaaacaaaaa caaaatgaca tcagtaactc 3240

ggacagttta aacaaaaccc ttataaccag ctgaatgatc tgtagaaatc cagttgtgga 3300

aaccggtagg tatgtgcacg agaggagagg ccattattcc aaggggtcac ctacagtgtg 3360

gaagaacaaa accctgatcc aacagtattt ctcaatgacc ttaagatccc cggggtcctg 3420

ggagagatgg gtcgcgttcc agcaggaggc tcctgagcgc cttcacttag gctggaattg 3480

ccgaggagcc atccacgctt aggcaaagct gcccgccaca cttcccagcg aaatcatcag 3540

tgcctgtaag agcagctgct ggaagagggg agcgagccgt ctgtcctgtg agcagcagga 3600

ccagggagac ggaggtacag ccctggcggg gggggcgtga cgccacgagc ctcggcgccg 3660

gccactgggc tttccagccg gaaggtttgg tacatctctg tctaccccta gggaaggagg 3720

tgagggagaa gttgagaggg cagcctggag ggggagggga ctctgctcct gcgtgtgttc 3780

tatcagggac tggctgtgag gtccggagct ctg 3813

<210> 99
<211> 960
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (716)..(716)
<223> a, c, g or t

<400> 99
atgctcgagc cggcgcatat gtgatggatc gcccgggcag gtaccttggg gactgaggac 60

cttttggctt ctctggagcc tgcaagcctc ttcccatgtg tccagctgct cttcctgcta 120

caaaggggac tgctcacagt ggcctcagct tggtggtttt gaggggccgc cccccggccc 180

tccataaggg tatcctgggc ctgagaattc tgcatctgcc attggaggat ggacagcctc 240

aaatggaagg agtcccacgg gagatgggtc cgaggtccgg ctgtggccat ccagcccccc 300

gtggcttgtc cagcctctgt gcacccctgg tgtcttcact ccaggggcag acagcagcca 360

ctgcagttcc tttcttcgtg agtaacagta gtgatagcag ctggggctaa caggctaggc 420

tttgtgttct gcgcatttgg tcagcttctc actcgatcct ccctaaagca atggggaggc 480

ccccactagc ccagttttca ggaagtcaac tgggaggtta gatgggggcc agggtcccac 540

atgctgctga tggcccgagc caggttgagc ttcctggtgt ccagtccgga tccccacttg 600

cagatctcat gctctcagat aggtgggaca agttcttttg tcacagtgct ggcctctgtc 660

ctgaggcctc attgctggct gggtgtgctc atgcgtggga aaaagcttgg gccgtnaact 720

atggccatac ctgttccctg tgtggacatt gttctcccgc tcccatcccc atcccgccgc 780

acccaaccca ccccaacacc cacaccccac gcccccaaac gcccacaccc acaacaccac 840

acccccccac ccccacccca ccacacccac cccaacaccc acaccctcgc ccccccccca 900

cctccaccca cccaccaacc cacaacacac cacacaccac ccaccaccag caccaccacc 960

<210> 100
<211> 2754
<212> DNA
<213> Homo sapien

<400> 100
gccagaagca gcctcagctt ggcaaggtgt ggagatgact gctgttccct tcgcatttgg 60

ggaaaacagg ctccctcggt agctcgatga tcctcttttg atcttgtgtg acctcctgga 120

gagtggatga cgctggtggc cttagctttt ctagacagtg taaattgcac tgggcgatgt 180

ccccagagca gggcaaggtc tctagagcgg gtctcccaca tgactggctt cacacaggca 240

cttccgctcg ggttgcatgc tctgtgtcat cttaccggtc cagggttgca ggtaggaaat 300

gtttgtaccc tcttctgatt gccacctcct tcccatcgcc ccttagggac agggcttgag 360

ggccagtgag gcgctggtca ggcaccccag gcctccttgg gacctgccca ggggcaccct 420

gagagctcct gaaaccccca cttagcttcc agacctttct gcaaaagctc ctcctggctt 480

tcctccctcc cccaatctat gggtcacagc taacagatct gagggcaact gctgtgctag 540

tggccagggc tgcacctgcc atccccggct ctgccacttt agggccttct agaggcagtg 600

tccttaggaa gtagctctga ggcatgggtt ttctgctcct gtgcagggca gctgatggga 660

taaggtgggg aaggacggtc agtgcttggg ccccagctgg ccagcctggc gatggggaaa 720

ccaaaccatg tcccccagcg aagggccaga gtgggaacct gtcctcatgc ccttcgtcct 780

gaggagccct gaggtgggca gcaggggcca ggggaagttt tcaggccttc atcaaagaga 840

acaacatcct cagctccgca cccctcatcc tgtatcagca cttaccggtg tgtgactgcc 900

cttgtcagct agcatacggt gggcccacct ggcccactgg ctgtttatgc cactgattta 960

tgatagggaa tattatcttt gaacccaatg aagtgttttc tcccccatca caaaaaaaaa 1020

aattcttatt tttagtagac atgtatttac caaaaatatg tactcaatta ttgtattttg 1080

gattttatca atttaaaaat tgtggaaatt tgtttgctct tacgccaaca taatattgat 1140

tttgcctctt ggctctgaaa gcccaaaata tttaccgtct agcccgttac agaaaaagtc 1200

tgctgactac tgagccagac ctccattacc tccatccctg ttggattatt taaagaaagc 1260

ctcagacagt aagggctttt ttaaaagaat aaaatgactt ggtttgcgct tggaagcagg 1320

ggaagcattc agatgagcgg tttctgcatt aaccctgcct atcacgcatc tcgtgtcctg 1380

tgtggctggc gagcccccct tggaaggttc tggtgcttca gctggctcct gcagagtcca 1440

ccccgcctcg tggtgggaat gcagagccct ttgctttcct tcttgccgcc tgcttcctgt 1500

tcctggggac ccgctgggcc tttggtctgc atcccctggc caggtccctc agggttgatg 1560

cgtggagaag gactttgagc agtggtgggc agcagtggcc tcctggccag ctcacactct 1620

tgtcctggga ggggcagcct gatctcacct ccacctagta ccttggggac tgaggacctt 1680

ttggcttctc tggagcctgc aagcctcttc ccatgtgtcc agctgctctt cctgctacaa 1740

aggggactgc tcacagtggc ctcagcttgg tggttttgag gggccgcccc ccggccctcc 1800

ataagggtat cctgggcctg agaattctgc atctgccatt ggaggatgga cagcctcaaa 1860

tggaaggagt cccacgggag atgggtccga ggtccggctg tggccatcca gccccctgtg 1920

gcttgtccag cctctgtgca cccctggtgt cttcactcca ggggcagaca gcagccactg 1980

cagttccttt cttcgtgagt aacagtagtg atagcagctg gggctaacag gctaggcttt 2040

gtgttctgcg catttggtca gcttctcact cgatcctccc taaagcaatg gggaggcccc 2100

cactagccca gttttcagga agtcaactgg gaggttagat gggggccagg gtcccacagc 2160

tactgatggc ccgagccagg ttgagcttcc tggtgtccag tccggatccc acttgcagat 2220

ctcatgctct cagataggtg ggacaagttc ttttgtcaca gtgctggctc tgtcctgagg 2280

cctcattgct ggctgggtgt gctctgctgg gaaaagcttt gcggggcttg cttggttaac 2340

cacagaagag aaggggactg tttggggtgc ctctctgcag cctccccgtg ctgggtggaa 2400

gcacggttac tgtgttctct aatgttcatg tatttaaaat gatttctttc taaagatgta 2460

acctccacac ctttctccag attgggtgac tcttttctaa aggtggtggg agtatctgtc 2520

ggggtggtgt ggcccttgga tgggtcaggt gggtgtgaga ggtcctgggg aggtgggcgt 2580

tgagctcaaa gttgtcctac tgccatgttt ttgtacctga aataaagcat attttgcact 2640

tgttactgta ccatagtgcg gacgagaagt ctgtatgtgg gatctgtgct tgggttagaa 2700

tgcaaataaa actcacattt gtaagaaaaa aaaaaaaaat aaaaagatgc ggcc 2754

<210> 101
<211> 301
<212> DNA
<213> Homo sapien

<400> 101
cagcgggggа atgtatagac aatatggccc atggtgcact aatgctgcga gcggccgccg 60

agatgtgatg gatggtcgcg gccgaggtac atttaatcca tcttctccat ttcctccttc 120

aggagccagc tatgagattt cagtgcattt ctagcccagc ggatgttcat tctccactaa 180

acttcatctt ttactaagca aagggggatt attccatgag gcagccagga gcaaggggcc 240

atgatattca tgactttgtc tgctgggcat ttttcaaagt gtccttgaat tcactcagca 300

a 301

<210> 102
<211> 4318
<212> DNA
<213> Homo sapien

<400> 102
gttgtatatg ctttttttтt ttccaaataa acttgtcacc ctgcatgccc ttggcaaata 60

agtgaagcag aaataggaac acagtccaca ttcaagttga ggaacagtgt atctttaaga 120

gctgaccttt gggtgacctg gaaaggggga aagatggcta agcatggaga gaaacgaggc 180

aagagacaag ctatgataca acaccgcttc agcccctgcc ctcaatagca cacaacccac 240

atatcagctt tctctagaga aggaacctac tgtttagtgc tcctcacttt gcaatgtttg 300

tgctacgcca gaatttctcc agtttttttc attatcatcc ccctgagaaa aaaattacat 360

tgaatttaaa ttttccctaa taagagaaat taaatatgaa agaataggat tttgttgggt 420

aagattgagc tttggaaggt cacgaaccat tattctatct aaggtgtgtg ttttgttttg 480

ttttttttg ttttttttg ttttttttt tgcaatccct ctccccctga accaatttca 540

cattgggaat gcaggaccta gactgctgaa taaaaagcta cttcttctat aattgtcagg 600

ttctctccaa tttctcagct tcctcaaaga catggggtgg agttgggtgt gccatcactg 660

agagagctct ggaacttttg actcttagtg acatttttag atttaggggt tcatggcctt 720

ccacatgttg ccaccaactg gtgatctctg ccccttcatg ctgatcaaga aagtagaaac 780

tcctcgtcgc cttcaggttt gcagtcgcag aaacattgcc tgctgtggac tgtcagcaca 840

aaactgggac actggtgtca tttagactgt cagcagtgca catgattgta cgatagactc 900

caggcaacca tgtgcatctg tgcaagatga cctctgcccc aagagaaggg ttacggtcta 960

attaaatgtt taccaagatg gtaccagtgg gctctccccg tgtcctttgg tgttattgga 1020

gctggtgtat gacagactca aggaaatttt ttaaggaaaa tgaagaagaa atcaaccttt 1080

atggttctct ttcattggaa gaggagaata aggaaagaaa tggcaggtag agagggaggg 1140

ggaaaggaat agaaatggca tgtctttgat gctgtggctt gtgtggggac aatgggaaga 1200

gcacagcagg cacaacatat ctgtgttagt gccacgtggt atctgttaag tatggccaga 1260

gcctcacata taagtgaaga gagttaagac aattcttgct ctttgaacaa taatagtcta 1320

tagaatttct attggcaaac catcccagac aacctgacct atcaaacaaa agcaaataat 1380

ccctgtctca gaactgctgg tctaaaagct agaaggggca tgataatgga aattgctaga 1440

aaagagagaa tgctctactc tctttcctgt gcctacctac ccccatccta aaccctgtaa 1500

aacagaattt caaaatagat gtcaaatatg aagtaattca gacttccaaa gaaggaaaga 1560

gttctgccca gggcagtatg agcaaatcca cagggatgtt aagatttggt ccaactcaaa 1620

ggtttatggg cagtgagcct agagtctctt gagagtaaac ccttgcattt gggacaagga 1680

gaatatgtga agttcaggag tgctcacact agagcaagat ccagaaaaaa aaaaatccaa 1740

tggcatttta aactagattg cattatcact caatgctgtt actttgagca gacaaatcag 1800

ttgaatggga gggcaaatgg cagaaatgaa caaaagctat taggtgaaag catccccaat 1860

gtatcagttg tgagatgatt tttgtttaat gatgatcagg tttacattga agtggcttgg 1920

aagacgtatt tcagagggac tgggtttgta ctgcaaaact ctgaacacta gagcaggttc 1980

tactagcact tgggcagtaa ggtagcgggt gtgtattatg ctattgccat agttctcgtc 2040

tttgtggatt cacataccct ttcccatgag gagcttgcta ctaacagcct cctgtttctg 2100

ttgtttttat atgggagaag agaaagagct tggaatttca attgtctaaa caattggtat 2160

gatttacaag aaggcaaacc attcaggaga tgttcaacag tgtatcattc ttagcattca 2220

atacaatgtt tattataaaa tatacacctg agtttatgtt tttctgccag gctgaagggc 2280

aataatgtct tgctatgcaa acactctatt ttatgtgtga attttttttaa ctgtaatttt 2340

atgtgtgaat ttttttaact aaactctatc atcattttct atgttgacat ctttttttt 2400

ttttaatctg tttccaactc tgagtctgtg aactatctct tctgactgga tgctggccta 2460

aaatcctatt agtgcttaaa cagacctcaa aacactctga acctgtaggc caatacagag 2520

atgttgttct ttgatatatc ttgggtcctt gatggcttca acaaacaagg tcataatttg 2580

tttaaattca gtgttttctg tagatacctt cttctcagtt tcattaagaa gagtaaccat 2640

ctttggtttt ccaaagaatg gaaactccta ccctagactt tagacttaga gcgtctgcct 2700

ttcacgatga gtaagagatc cagaatgttt aggggatgtg cgtcagccca cctggtacag 2760

tggcaaagct gaatgcgaac ttgaggcctc ctgaggcctg accttctatg cccccagccc 2820

tccaccccaa gcacctttg gtgtggctgg aaacacatga cattggatgt gatttttctc 2880

aagcccttca ctgtggaagg catgggagac tgcccagcat tgggcatgtg gctgttaacg 2940

tttccatttc aagtccctga ttcttactgg agaagttaag gagccactta atgttttcac 3000

agacctctga ttctgttgtg aatgtggcat tccagtgagt acaacctgct tgctataaag 3060

aaagcagcat attttgacaa ttttatttct tccttgggta cttacatttt tattacaaaa 3120

atggccgtta taaaaaaaga cagaagggct gggcagtggc ttgctcccta gaaactgaga 3180

ttccgaagca ggtgtttcct ctcccctaga ctcagaggta catttaatcc atcttctcca 3240

tttcctcctt caggaccagc tatgagattc agtgcatttc tagcccagcg gatgttcatt 3300

ctccactaaa cttcatcttt tactaagcaa agggggatta ttccatgagg cagccaggag 3360

caaggggcca tgatattcat gactttgtct gctgggcatt tttcaaagtg tccttgaatt 3420

cattcagcaa acaaagcttt ctggagaatc tctcaaaact taggccctgc tccatttggc 3480

caaaaatgat ggctgctcca aaactctgaa cttcttaaaa ctccatctgc tacattattt 3540

ctggagttta aacatgactt tttctgtctt tgagtataga tgtgtttgtt taattaacga 3600

agcacaagtc tgttaagcag aaggctccaa gctgtattct atacttggga atccttggtg 3660

ccatctttat tctaccaagt gccaatcacc atggctaaag tgggcgtata ttacagcctg 3720

tgtcctaagc ttagaagctt taatgtactt ttttaaatga aaagtattag agggggttga 3780

acattgtaac taaagcataa agttagacca attacatgca gagatgttta tttaatattg 3840

tgtgagctga gtccttctgt ataaattatt tgcacacttt tcttgcatga tgaactgatt 3900

ttttatagtt gtttgtacca gacggtggca tatttttgta aaaaactttt gacactgaat 3960

tgcaataaaa tgtttttcca acaaaaaaaa aaaaaaaaaa tgcggcagag gctgcggggc 4020

gacgcgcggg ccggcgcagc catggtgaag attagcttcc agcccgccgt ggctggcatc 4080

aagggcgaca aggctgacaa ggcgtcggcg tcggcccctg cgccggcctc ggccaccgag 4140

atcctgctga cgccggctag ggaggagcag cccccacaac atcgatccaa gaggggggc 4200

tcagtgggcg gcgtgtgcta cctgtcgatg ggcatggtcg tgctgctcat gggcctcgtg 4260

ttcgcctctg tctacatcta cagatacttc ttccttgcgc agctggcccg agatgcgg 4318

<210> 103
<211> 2288
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (948)..(949)
<223> a, c, g or t

<400> 103
tgtgacggaa acattcacac ggaacagcta ggccatgata cgccagcttc cccgtctatg 60

ggataccaaa gttgccagca attaatttac attaccaaaa aggagaaacc attatttatg 120

ctatcaccaa agccataccc ttaaaagtaa aaacacatct ccacttgata ttagtaattt 180

gagaaaactt agcatatcta caactactgc ttaatccaca acttttcatt gattcaacat 240

tatcacttaa atgcagatgc aaaataaccc ttgcatggca aacatgttta ctctgtgcta 300

atgaaactat gaaacagttc ctgcaagttc acagacacag tcaatcaatg attgactaca 360

gaacagactg gtgagtgtta ccaaaaacaa aagacctgcc caggggacca ctaggggtgg 420

aggaagaaca cctgggtctc actcactggg cacttgatgc tttcattttt atctgtaagc 480

caggggacaa ttatattaat attgtatgaa tacactgcta ctgggaaaag cattcattca 540

acagatgccc agctacatac tctgtaccag gtagtgtgtt aagtgctgaa acagggaagg 600

acaaggcaca tcctacgggg gctcgcagtg tgggcaggtg ctgagctaac gcacacacac 660

aagggctggg tacctgccag atctgtatca ctccggagcc aacgacctgt ctctctgaat 720

gatctctggc atggtaaaag agttaactgc tatttcaaac aggctgaaca caaaactttt 780

tttgccttcg cttactgcac tgcttctcaa acttgactca cgctggcttc acggcctgcg 840

acgtactgca ctcacgaact tactcctgct gaacgccttg cgccgctaac tccactgcgc 900

tcactacgcc tgcttctccc tgctgctgca cacactctgc tctactcnnc cgctcacaat 960

ctctccacac aacataccgg agccagcccc ccccccctcc ccccgccccc cccctgccg 1020

ccccccccc ccccccccc cccccccaac actataactc ctgcgctcag cgcctcacct 1080

gcccccgcac tccccccagc tccggcccac acaccatgcc ccgcagccca accctcgcca 1140

ctcccacact gcacctccgc gcctccaaac agcccccgcg accaacacac accgcgccca 1200

cacgccgact acccccggcc gcccagccca cccaccactc ccgacatgcc ccctccgccg 1260

ctccaccata gaacacctcc gccccggccc cgcgcccgac ggccccgaca ccgcgccccc 1320

gccgccgcca acgcccgcac gccaccaccg ccacccaccc ccacacacgc acccagccac 1380

cactacccgc caccatcccc accccaccac tcccacccgg cccacaacca ccgcccggcc 1440

acccaccaac acacgcccgc acgaccccca ccccgccccg ccccaccacc tcgccccgcc 1500

aacaaccgcc acccacaacg cgcccgcccc accgccccag ccacccatcc acacccaact 1560

accagccccc gccccccccc ccccgccacc gcaccccact ccaccaagac gaaaactccc 1620

gacaccgccc actaaacgca ttacagcaag ccgctgagac ccgaaaacaa ccccccgcaa 1680

gcagaggcac cgctaccatc aaacgcaaca accacacaac gcacggctat catccccccc 1740

ccaccgtgac cacaaacgca ccaccccgcc cgaccagcac gccacacccc cacggccatc 1800

acgcacagca ccacacccca cgcactcaca cgacgaaata cccccacgcc agccgcccaa 1860

ctcccacccc ctacggcaca aacccacacg gcggacccac acctaccaca cacacgcaac 1920

acaggacgca tcagaactac ctaaccacac gcacgaacac acaacagcac accacccccc 1980

aaccacggac gcacgagctg tcgacccact accaccacac acgatgaccc gacacagcgt 2040

gcgagcgcca tacccgcgga gcagcacagc caatgcacgg ccccacgacg gcgacacgcc 2100

gccatccttc agcctcgcgc caccccccgtc caccgaacca agaccaagac gcagaatcgc 2160

aggccgcacc acacccgcta cggtccgtag cgagacagca aacaacaaga cgactgcggc 2220

gccgccacca ccgccaagca actacgcaca accaccgacg cctcatcacc ccaccaccac 2280

aacactta 2288

<210> 104
<211> 592

<212> DNA
<213> Homo sapien

<400> 104
ggatatcgac taactatagg ccatgttgac taatgctgct cgagccggcg ctgtgtgatg 60

gattggtcgc ggccgaggta cacatattca tgccactcgg ctctgaaaag aggttcaagg 120

tgggtcaagg tgggtcttgg ccagtggagg aaggaaggtg tccaggactt taggttaatc 180

aacagtgggg acagagagga atgatttccc ttgtgaaaac aacagggttc ctttctcata 240

ttcttgtggc cagaaactgg ggtgaacttc agttgtggtg taattgaaat gaaacagtga 300

gagccatttc tccatggaac tcctatgacc tccattttaa cttctgacaa agttaacttc 360

atttatacaa tctgtatttg aaaacagtaa tcacaaccaa aaaggtccta taaacctgta 420

atagatgtca aagggattca cattctgaac tttaatttta aggacccttt aaaaggccta 480

gacttggatt aaagtaaacg taatattcca agctaaaacg aggcaccata aaaaatcaac 540

tcaaaacatc caaacaatgg ctagatggac taatgtaggt tgttttgctt tt 592

<210> 105
<211> 2180
<212> DNA
<213> Homo sapien

<400> 105
ttttttttttt acatctgtac aagatttatt tcatttacca caagttttag agtgaaagta 60

tatgccctaa tacagaagga gtgattcaga acccaccaaa agatcaagag ccaaagtcag 120

ggcttgtatc ccaggtcaat aggagtttca tccgccattt gtttttctgg gattaaagag 180

atgatgtgca gaacgcccgg gctgagctca ctatggcttc tgacatacaa gcaccactag 240

gaacaagact tcatgcttca tccttgctct ggctgtgcta gtgagtggca aaagcctctg 300

ctcgtgcttg ccctccaaat ttttcacaag cattcctgtt actagccata gcttcccatg 360

acttcacatt tgagtaagtt atcctgaagt gtgtggctgt aatattgaca aatagaacta 420

atgagagtga actgccatct gtttaacaaa acgcactctt ctaagagaat acccagtctt 480

agtgcattta gtacatagtc acaacattga aatgacgtga gaatcaatac aaaatattta 540

tttttttttca aaccacagaa ttcttaaccc cagagccaca caataaagtt ctcagaattg 600

taagccatta acatttttct aaacaatgca gtccagagat gaagataatt tccaaccagc 660

agggatgcaa tatatagtag gttcccctat gaatgaagct caaattagca tttcctttaa 720

ttctcccaca gccactccat caacagaagc agaaacagta cacatattca tgccactcgg 780

ctctgaaaag aggttcaagg tgggtcaagg tgggtcttgg ccagtggagg aaggaaggtg 840

tccaggactt tagttaatca acagtgggga cagagaggaa tgatttccct tggaaaacaa 900

cagggttcct ttctcatatt cttgtggcca gaaactgggg tgaacttcag tggggtaatg 960

aaagaaacag gagagccatt tctccaggaa ctcctatgac ctccatttta acttctgaca 1020

aagttaactt catttataca atcgtattga aaacagtaat cacaaccaaa aaggtcctat 1080

aaacctgtaa tagatgtcaa agggattcac attctgaact ttaattttaa ggacccttta 1140

aaaggcctag acttggatta aagtaaacgt aatattccaa gctaaaagag gcaccataaa 1200

aaatcaactc aaaacatcca aacaatggct agatgactaa tgtaggttgt tttgcttttt 1260

agttgcaaag cttttcagta tctcagatta gtgtatgttc ataaaacaat gctcagttat 1320

tttaatagct gcttatgaaa caataacagt ttaactcaag ggcaatgctt cttgcataat 1380

aatcacaaaa ataattaact gctataaaac gggaaaaaaa gtagaagaaa ataaagccag 1440

cctcaataat aaaaaggcaa aatctggagg ggttactcgg cttaaaaaga gattaaccag 1500

gattatttaa atactataat acacagtgct cctgctcact tctaactgca gaaaccaatt 1560

ttgtttgcta gatcaccatt acctttgcta gtatgcgtac agaccaccac tcggaagtta 1620

tccttttgtg ctgaaaaacg ttcaaatccc ttgtttggtc agtacagaat attgcgaggt 1680

gatgctcatg caaactcttc ctgaggaatt tatgtgtgca aatctgcaac ccgacagcat 1740

ggcacgcagc cagggagtgg tagctgcaca gtgtgagcac tggagatgga tgtgcagtgt 1800

gcagtgttca cagccatgga catccattct tctgcaatct catctaccca caaattagct 1860

ttcactctag accccaaagg gagggtaatt gctgcaaatt tgttaaaggg acagaagaaa 1920

aagtagcttg tctacaaaat aatgcacaat gcatgcatct gggtttgtgt ttcttctcac 1980

taaccttgcc taagaaccat taggataaaa gtcacaacac caggttttgc tttctgcccc 2040

acaaaaaaac agtagttaat tcctgtcagg ttagggtaca ggtgtgacaa caaaaggtca 2100

caaaatgaca atgttactga agcttaaggc caacctttaa aacatgtacc gtctctcaaa 2160

acaattatcg atttactttt 2180

<210> 106
<211> 611
<212> DNA
<213> Homo sapien

<400> 106
cgggcgtctg gcgcaggtac gttacggagc tatctgggag gaggaagccg tccacccaag 60

ctcacccgac tggttcaagg ttaacaatgg agagcctcac aaacacacgg aggcacttgc 120

gagaggcaga ggctggggcc acagagctgt gaggagagag tgaatctccg tgcatggcag 180

acatgttgtg agaggcagag gctggggccg cagagctgtg aggagacagt gaatctccgt 240

gcacggcaga cacacaaacc ccatcctcag ctgcagattc acaggagagc atgggggcaa 300

gtgggctctg cagttataca gcaggacagc gcacggcctt gggcaggcca caatttctta 360

aagggaatgg gaaaatgaga gcttacggaa agcaggataa agcgccctcc actcatgctg 420

aaaagcacga agacctgctt tccaggaatg acctgcctga ggccgggaca gtcagctacc 480

aatgcaagac acaagagcaa gaaggcccaa gtgcagagag cgcaagttaa ctcctccagc 540

tccaagacac agggtgagag caggctacgt gtgggcagag acggcatcta tctgcccatc 600

tctgggaaaa g 611

<210> 107
<211> 1845
<212> DNA
<213> Homo sapien

<400> 107
cacgaccatc tcgcactacc ctgactacag ctggagcgcc cagaggcgag ccctggtgga 60

ggaagaggtg agctgagtgg ggcctatgct gggggaagct ggcccttccc acattccacg 120

ccctgagtgc ggagtgctgt tttccgccgg ggtgctttga gtctcgctct gtcaccaggc 180

tggagtgcag tggtgcaatc ccggctcact gcaacctccg cctcctgggt tcaagcgatt 240

cttctgcctc tgcctcccga gtagctgggg ttacaggcac acaccaccac gcccagctaa 300

tttttgtgtt tttggtaggg acggggtttc accgtgttgg ccaggatggt ctctatctct 360

tgacgtcatt tgtgatccac ccgcgtcagc ctcccaaagt gctgggatta caggcgtgag 420

caatgtaata aacagacttt atttttttaga gcagttgaag gtaacaggaa gtggagctga 480

aaacacaggg tggccccatc catgccctct cacagcccct ccttcctctg agcccctcac 540

gtccttggct ctcctccgaa gcttcttttc ccagagatgg gcagatagat gccgtctctg 600

cccacacgta gcctgctctc accctgtgtc ttggagctgg aggagttaac ttgcgctctc 660

tgcacttggg ccttcttgct cttgtgtctt gcattggtag ctgactgtcc cggcctcagg 720

caggtcattc ctggaaagca ggtcttcgtg cttttcagca tgagtggagg gcgctttatc 780

ctgctttccg taagctctca ttttcccatt ccctttaaga aattgtggcc tgcccaaggc 840

cgtgcgctgt cctgctgtat aactgcagag cccacttgcc ccctgctctc ctgtgaatct 900

gcagctgagg atggggtttg tgtgtctgcc gtgcacggag attcactgtc tcctcacagc 960

tctgcggccc cagcctctgc ctctcacaac atgtctgcca tgcacggaga ttcactctct 1020

cctcacagct ctgtggcccc agcctctgcc tctcgcaagt gcctccgtgt gtttgtgagg 1080

ctctccattg ttaaccttga accagtcggg tgagcttggg tggacggctt cctcctccca 1140

gataccctca cgtaccctgg ggttctgggt gtcacagaag ctgctctggg tgcggccact 1200

ctggagcggg gcccttgtgc caggtgagac acccttctcc tcagcaagcg gacttggtgc 1260

gcagcaggcc aatacgtcgg tgccgacatg ccatgaagca cgggaggccg taggagattg 1320

agggaggtga atgactcagt cctcatcgta aggagcgcgc tgaggccgaa gatgagcagc 1380

ggagctggtg catggggccg actggcccca tccatgccct ctcacagccc ctccttcctc 1440

tgagcccctc acgtccttgg ctctcctccg aagcttcttt tcccagagat ggggcagata 1500

gatgccgtct ctgcccacac gtagcctgct ctcaccctgt gtcttggagc tggaggagtt 1560

aacttgcgct ctctgcactt gggccttctt gctcttgtgt cttgcattgg tagctgactg 1620

tcccggcctc aggcaggtca ttcctggaaa gcaggtcttc gtgcttttca gcatgagtgg 1680

agggcgcttt atcctgcttt ccgtaagctc tcattttccc attcccttta agaaattgtg 1740

gcctgcccaa ggccgtgcgc tgtcctgctg tgtaactgca gagcccactt gccccctgct 1800

ctcctgtgaa tctgcagctg aggatggggt ttgtgtgtct accgt 1845

<210> 108
<211> 160
<212> DNA
<213> Homo sapien

<400> 108
ctgttacttc cttttgctgt cagagggttg ttgacaatgg caaggggga tgttagtgag 60

atccaggtgg tggtggcttc ttggtccact cagcttgcac acatgcaaga ggagggtctg 120

tggcccctca gccgtgctgg tggcctcctt cctcaagctt 160

<210> 109
<211> 4621
<212> DNA
<213> Homo sapien

<400> 109
ctgagagaag cacgggggga ttgattggtg cttttagaga acacacctag atacaaatgt 60

gatacatttc tatactcaaa ttagaacaca tgataatgta gatttgttca ttcatcaaag 120

acaatcaccc attatagttg taattctcag tatttggctt ctcttgtgga tgatcagtat 180

tacttttctt tcataggtat tggatatgct ttttaggtct atcaaaaact aaagtctctg 240

aagtgacact ggtacaagca attcacttga atttctaaaa acagataaaa tataatggac 300

cggccgcggc cgccgggagt tttttatttg gaatattatt agatgatgcc ctatgataag 360

atgagacaaa tgatggggag ggaaggagga tggcccttct ctagaatcag caaacccaat 420

ggttctgtga aggtcagacc caagcttgga gcagctggtc tggatgcaga tccaggagct 480

gcagagtgtg gaacaggaaa ggctctgccc agggcccacg agctctagca ttcctgctgg 540

cagattagag atcttagtta aatttgacca agaaaggagc ttagctaaga ggtctttgtt 600

tccagaggac ccaagactac cagactcctg tgcagtctgc ctgtctgcag agcctccaat 660

cacggtttaa agtggtcttg gcctcccagc ccccactgct gcagccttca gagccagcct 720

ctggactcag cgccaggagc ctcctctcct gctgccctgg tacctgtcat cctagtgtca 780

cctatttgtt ttgcacagca ttctggagaa catgtgcctc acaaagtgat cctgtttttc 840

tcaccctcca aagctgaggt ggctaaggga ggcaggggtg agagagctca ctctctaaac 900

cacaggcagg gcaaggtgct tctaggccgt gtcccctccc tggggaatgg tcactactcc 960

tcagcatggc cacaggcccc tgtggttttg ggagtttgct gagtgactgg aatgggctga 1020

ccccagccgt ttccctgctg tggaagcagc aaggtagggg cactgctcgg tggtttgtca 1080

agctcagagg cagccacgga tggttaggaa caggaagcca cactgctaag gtgtgatgag 1140

ctagaaggga gacatcctgg acttcaagtg agacaagtgc agaggaggcc actccactta 1200

ggaccttcct cttagacctg catgagcacg tgtttcactt aagcacattg ggtccgaatt 1260

gtggcgcatg aacccagctg ttaaaccagg taccctggga ctgccctctg cctctcagag 1320

ccctacctgt cctggacttg aaggcgagca tcaggcaggc cttctcagcc catcttccac 1380

caggagtgca gagccagggt gctgagagga gcagggactg taactacagt aggaggcaac 1440

tttgctggtg tcattcctgc tcggattgac tttgtgctcc ctcccatcca cagctatggt 1500

agccagtatt tccccagcag ctctctagtg gggaagagtg gctagaacca ggccattctc 1560

agtcgaattt aggaagacca agactcatgt tcctgacagg acggtcattc cctcatggca 1620

tcactgactc agttctaatt cctcccctcc tgtttcattt tggtttgaac aacctgtgcc 1680

tgcagctgtt tcttccacct tcctggagct cacaattgga gcttcatgtt cagtcttgag 1740

ctcctaatgt ttccagacag ggcaggggat aggacagcca taatctatga aagctgtgaa 1800

acaaaccctg ggttatatag acctcagatg gtctcatctg aaggctttgc ccacttccca 1860

gtgagcagag ggtttgggga acaaagaggt agctgaagag tggtggtttt tccctgtatc 1920

ctgattttca gaaagggctt tagtctctga gctgtggcta gtagctggct ggctgaccag 1980

gggacggggg tctaataggg tgtctgacca taggacacct ctcgtgtttg gccttccctt 2040

tcttgtgtgt ttgtgtgtgt gtgtgcatgt acacacacaa atctgtatcc ccaggagtgc 2100

ctcctgcagt cactgggggc tcttccgtgg tgcaccatcc catgagagca gatcagaggg 2160

ctgtggtcag gacttcctcc cgagctgccc tccctccctg ggatgagttt cctctccacc 2220

tgctccaagg tgccttacga ggtgcagctc ttcacctggc cacatggcat ccatcaccgt 2280

gttgagaaat ggtgggacgt tggctggata ttgtgtgcta tgaatctgat ttgtccaaat 2340

ctggacctca gccaagccac ctgtggtgga accttgcaag tcaaggtcat gaacaaattt 2400

cttgtaacta ttggaaataa actaatgagg tagaactgtc atatccaaat ggaagaggtg 2460

gttggtactg ttacttcctt ttgctgtcag agggttgttg acaatggcaa ggggggatgt 2520

tagtgagatc caggtggtgg tggcttcttg gtccactcag cttgcacaca tgcaagagga 2580

gggtctgtgg cccctcagcc gtgctggtgg cctccttcct caagctttaa cttgtgccct 2640

ccctggcaac agggagtcat gggttaatgt cacctgtcat gcctctttct tcagggaaaa 2700

tattaaaagg ctcaaatcca agctgtgact tcccttagac agtgtaagcc tgtagggtgc 2760

agccccaagt ttatgtccaa aggctgcctc agaaccctgg agggagcagc gcccagctgg 2820

cagctcactc tcccccttgca gttcactggc cttttcgcag ccaggctaag ataggcagtg 2880

caaggagctt cccactggcc agagcagctg ccttgtagaa tgtggggctg gcttttagaa 2940

gtttcacacg cttgcagcta agagtgcaga gctctctgtt ttacttcgtg aaacattagc 3000

agctctgagg ggcttggctt ttggtaaggg acttatagtc tgacttgcca caagcacctt 3060

gagctcctgg gttggttggt ggttggttgg ttgggttgtt ttaaatttga ttttttctttt 3120

cctgaagttt gttagaagtc agacctatgc aaggaaaacc atggcccact tcagcagcca 3180

tccagtgggg gtgcttaatc ggtggtggta ggattcatgg aaacgttttt ttgaaatgaa 3240

agagaagatg atgtacttac attgtaaaat ggtttacaat aaagtttggc atcttattac 3300

agtttttttt aaaaagaaac catcacattt gtggttgttg gttgtgggtt ctctggagtt 3360

gaaacagctt ggcagggtca aaaaagattc ccctgtgaat tagaaggttc tccccaggga 3420

ccagagtgag aaggaacctt acgaagcctc atggcaggag tgggagccag caactgccca 3480

gtgaaaagcc aaacatgctc ttgccatgtc catctccact tcttcgctag tgagaattaa 3540

agcaagtccg tctctgagga ggacatcgtg ttggggaggt agagtataag cagtgagtag 3600

cccacatcct tgtccctgag gctgcggggc tggaaggagg tccctgggca gagccgcagc 3660

atcaggtgcc atctgcaggg gcacccctgg aactagtgga aggcccccac aaggcctggc 3720

ctgccctcag ctgcctcttg gcccatgtgc cacatcctgg gcattcctca tgtcagcact 3780

tgtctcagat caagcctggc agtcctctca ttggcctcct ggagccaggg tgtcaggagg 3840

gggattgggg tggggttcaa agtatgacag caatggagtg gcatttggga taaccaaatg 3900

tcagtgcctg gttgagttcc atctgcctgg agccttcgag gataaaaata atctccatca 3960

taaactgaaa aggacttaga tatgtgtgtt actgggaggg ctgcgaccta tgcacaccta 4020

ttccccatac ccaccccagg cgtaggctcc tggcctgtgc cctgggggat cttccccttg 4080

caaggagggg tgtgttagga gcttctagca aaggagagct aagtgacctt ccggggagcc 4140

cttgggcctg ttcttgtctc tgcagctcag tatctggagt gggcttcact ctcccccacc 4200

ctctttgagc aactgaaact gtcctgcagg ccagcccccc tgcggtggcc tcacctactc 4260

ggaggtgagt aggctcaggc ctgagcctga gcctgtttcc cacctctgtc tcatcagtgc 4320

tgtcaccatc gagcatggtt tcctccctgg tcctttccag acccccttact ccccccctctc 4380

tgaacttgca gttggaatag gatagaggac aaggacatca gcaaaaacat gagcaactcc 4440

ccaggctccc tttacctcac tttctccacc tgtgaacctg tggtagcccc tctctgagtt 4500

tgctgcaagg gcccagtgag cttgccataa tcagacgcag cagacacggg aagctcctgc 4560

tctgccacca tgcagggatg gaaaatctgt ccactggaat ccaggggaag ggatggaaga 4620

g 4621

<210> 110
<211> 303
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (176)..(255)
<223> a, c, g or t

<400> 110
tccctgttat tgttccgtat attacctgta agcagatact gtattttatt ttagcctatt 60

tgacagaaca catcactcag aaaaagtgaa gtttcagagc aaacagtgaa gaaatcagtg 120

tgattgtaga caaaaagtcg gttcacagaa cggagcagcg gggggaggaa gggaannnnn 180

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnnnnn nnnnncccgg gggcagccgg ggctatgagg ggacgtgctg atgcgtgatg 300

atg 303

<210> 111
<211> 1848
<212> DNA
<213> Homo sapien

<400> 111
tgtttttttg agacacgggt ctctggttaa ccccggctgg ggtgccatgg tgccatctcg 60

gctcactggc aacctttgcc tccccgattt aaggcggttt tttcccgcct ctagccctcc 120

ttgagtatgc ttgggactac agggcgccca cctctcacgc ccggcttatc tctctgtaat 180

tcttagtaga gacaggcttt tcacccacgg tggcccggct ggtctccaac tcctggcctc 240

cagtggtccg ccccccttg ccttcccaag tgctgggatt accggggtgg ggcccccgtg 300

cctggcccgt cctatcctcc cccccaccct ttttttttt tttttgacat ggagtctcac 360

gccatcaccc aggctggagt gcagtggcgc catcttggct cactgcaacc tccgcctccc 420

atgttcaagt gatcctcctg cctcagcctc ctgagtagct gggactacag gcgcccgcca 480

ccacgcccag ctatttttt gtatttttag tagagacggg gtttcgccat gttagccagg 540

atggtctcga tctcctgacc ttgtgatcca cccgccttgg tctcccaaag tgctgggatt 600

acaggcctga gccactgtgc ccagccttat cctcaatttt ctaagcataa ggatttaaga 660

gttagtggga aagcatgact acggtttaaa agaacctggt cagctgaagc agtcccactt 720

ggccattcca gtacacttcc tactctcgct tcatggagaa gaccgtctac tgacaagaga 780

atccatcccc tttgcacaca gtggtcacct gacagctcag aggtcaactt cttggttttg 840

aggaaaaatt caaaaggctg cacgtttaaa ggagttgttc catttcagtg ccccattagc 900

ctgtctttcc tcagaaattt accaatctct tgatgtgata agcaccaaac tatgaagctt 960

ttcccttcct ctccccgctg ctccgttctg tgaaccgact ttttgtctac aatcacactg 1020

atttcttcac tgtttgctct gaaacttcac tttttctgag tgatgtgttc tgtcaaatag 1080

gctaaaataa aatacagtat ctgcttacag gtaatatacg gaacaataac aagggactga 1140

aggtacaaaa cacacacgga cattatggtc atataaaata tatagaagta cctaacgtta 1200

acaatttttg taagaagttg cacattttgc tatttttta aagcatttcc ccacattttt 1260

tttcatctgt agatcctctg cagttaagat tgagcacccc cccgcaactt tttttttga 1320

tacagagcct cgctctgtca ctcaaggctg gagtgccagt ggcatgatct tagctcactg 1380

caacctccgc ctcgcaggtt caagccattc tcctgcctca gcctcccgag tagctgggac 1440

tacaggcgtg cgccactgtg cccagctaat tttttttgt atttttagta gagacaaggt 1500

ttcaccatgt tggccaggct ggtctcgaac tactgacctc gtgatccgcc tgcctcggcc 1560

tcccaaagtg ctgcgattac aggcgtgagc cactgcgccc ggccaagatt gagctccctc 1620

ttgcaatgtt tctcaaattc atcacgattg gaacataccg acggaattca ttttctgtct 1680

gaagcaatga gtgctataca tgaaagtttt cctcacatat aatattcatc aagtttattt 1740

acagtataag ctttccgatg ctaagggctg agatgtgatt aaaggctttg ccgtattcat 1800

cacattcagg aggctggggg ggggggatta atttaattaa ggatcctt 1848

<210> 112
<211> 333
<212> DNA
<213> Homo sapien

<400> 112
ggccgaggtc tagaatgggt gaaggaatgt gttagtgtct ggagccatga aggccttgtg 60

ctgaccaagc tgctcacctc ggaggagctg gctctgtgtg gctccaggct gctggtcttg 120

ggctccttcc tgcttctctt ctgtggcctt ctctgctgtg tcactgctat gtgcttccac 180

cctgcgccgg gagtcccact ggtctagaac ccggctctga gggcactggc ctagttcccg 240

acttgtttct caggtgtgaa tcaacttctt gggccttggc tctgagttgg aaaaggtttt 300

agaaaaagtg aagagctgga atgtgggggа aaa 333

<210> 113
<211> 2269
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (2266)..(2266)
<223> a, c, g or t

<400> 113
gcgagcggct ggcggatccg acgcgcgaga ccgggagggg acgagggcgt tgcaatcgtt 60

cggggcgggg gctttccggg gagggggtgc tcaggtgcac cagcggcggc ggaccctcag 120

actctgccct cccctccctt taacccccct ccagccggac gggaggcggg gcagggctga 180

gcatttgtga cacctacatt tccgtggctc ccttcttttc ccccgacccc tgtttatctc 240

ttcgccttcc agaagttctt ttccatcagg ccgtcgcacc ttgcgtggga aggagcaccc 300

cacttggaag caggaggcgg ggttcagatc ttggccctac ccctcctgtg ttaaagtccg 360

cgagcctcag tttccctcac agtatttttt gcctcgcctt acccggtttt gaggatctgt 420

acgagaaaga gaaaggaagt ggacatttgt tgaattcctg catggccaaa taccacgcag 480

actgcttcat ccgccacgtt taatccttat tacttggtgt tctcagaact cccatttcat 540

ggattcttaa gctcacagag tcagtgaata acagaaaggg attcagatct agccgtttag 600

ctgcacagtg gagttcttct ccagagtctt cccttgtctg ggctctggct ggaactattc 660

ctcagccaaa tcctcgcccc agaacagtgc ttcctgtttc tccagctgag aagtctccct 720

ttcagtttcc ttcttccagc acggagtaca ctgctctgcc tccacttaga ttacttcaga 780

aatgaaatgc agcaaatatt tatccagcag tgcagggagt tgaacttttg gagtcgggaa 840

ccttggattc ttgttctggc tctgccactt actgtgtggc cttgggaagt cctttgtctt 900

ctctgagctt tcttttctct ttgcgtaaaa gcggtgctct tgtcccattc tccctccctg 960

tcttccagca ggctctcccc ggaggctcag ccccctctgc tccccatggg caactgccag 1020

gcagggcaca acctgcacct gtgtctggcc caccacccac ctctggtctg tgccactttg 1080

atcctgctgc tccttggcct ctctggcctg ggccttggca gcttcctcct cacccacagg 1140

actggcctgc gcagccctga catcccccag gactgggtct ctttttgag atctttggc 1200

cagctgaccc tgtgtcccag gaatgggaca gtcacaggga agtggcgagg gtctcacgtc 1260

gtgggcttgc tgaccacctt gaacttcgga gacggtccag acaggaacaa gacccggaca 1320

ttccaggcca cagtcctggg aagtcagatg ggattgaaag gatcttctgc aggacaactg 1380

gtccttatca cagccagggt gaccacagaa aggactgcag gaacctgcct atattttagt 1440

gctgttccag gaatcctacc ctccagccag ccacccatat cctgctcaga ggaggggggct 1500

ggaaatgcca ccctgagccc tagaatgggt gaggaatgtg ttagtgtctg gagccatgaa 1560

ggccttgtgc tgaccaagct gctcacctcg gaggagctgg ctctgtgtgg ctccaggctg 1620

ctggtcttgg gctccttcct gcttctcttc tgtggccttc tctgctgtgt cactgctatg 1680

tgcttccacc cgcgccggga gtcccactgg tctagaaccc ggctctgagg gcactggcct 1740

agttcccgac ttgtttctca ggtgtgaatc aacttcttgg gccttggctc tgagttggaa 1800

aaggttttag aaaaagtgaa gagctggaat gtgggggaaa ataaaaagct tttttgccca 1860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1920

aaaaaaaaaa caagcaaaaa gacaagagaa gaaaacacaa gaagaaaaaa aaaccacaaa 1980

caaagacaaa agaaaagaaa agaagagacg aaaagagaga aacaagaaag aggagaggac 2040

aaagagcaag aacgaagaga gaagaaaaga gagcacacaa agagaagaga agaagcgaga 2100

ggaagcgaaa aagagagagg aaaaagagaa aagaagagaa gaccaaaccc gagaagaaga 2160

gagaggaaag gcgagcgagc aaagagcacg agaagagaag aaaggaagcg agggtgccaa 2220

gagaagcgga gagaacgaga gggagaaaag agggagagga agaagnagt 2269

<210> 114
<211> 550
<212> DNA
<213> Homo sapien

<400> 114
agagatatat agcccgtctg ggctaacttg tgtcctctag atcatgcgtc gagcggtcgc 60

cagtgttatg tatcgctggt cgcggccgag gtacacacaa gaagccagaa gatatttttt 120

tttcagtgaa ctttctcctg gaagcaaagg agaagctatg ggagatccag gcatggtttt 180

ggcttctgga ggctgttttt tggttactgg ggtctcttca aagcaaaacg ggatcaggat 240

gaagaggggg aaaggcatgg gccataaata aataaggaaa tttgccccga ttctacaaat 300

gcatctgatg gaatgaggag ggtcaggctg tgagggattg ggggtagctc caggctgctg 360

ttagggacac acttgagctg gttcattccc tggagtgccc cgcaactgcg tctagggaga 420

ggcaggccca agcacatgtt cagtctaatt ccagctcaca ctggaaggct tcctataggt 480

ttggttgccc tgatcacagc cgagctgccc caccaggaat taaccctggt gggcaaaggg 540

tccatcttct 550

<210> 115
<211> 4441
<212> DNA
<213> Homo sapien

<400> 115
agcgggggca tcgccgcccg cgcccctcta agtgccgggc cgcaagctcc accgcagccg 60

cctgcaagca gcggcgcctc ggccctcgac ctgcgcgcaa agcctgtgct ggagccgtcc 120

tcccgcggcg gggaccggga ccggggaccc aagccaatcg aaagctccaa ccatggccat 180

ggggctcttc cgcgtgtgtc tggtggtggt gacggccatc atcaaccacc cgctgctgtt 240

cccgcgggag aacgccacag tccccgagaa cgaggaggag atcatccgca agatgcaggc 300

gcaccaggag aagctgcagc tggagcagtt gcgcctggag gaggaggtgg ctcggctggc 360

ggccgaaaag gaggcactgg agcaggtggc ggaggagggc aggcagcaga acgagacacg 420

cgtggcctgg gacctctgga gcaccctctg catgatcctc ttcctgatga tcgaggtgtg 480

gcggcaggac caccaggagg ggccctcacc tgagtgcctg ggcggtgagg aggatgagct 540

gcctgggctg gggggcgccc ccttgcaggg cctcaccctg cccaacaagg ccacgcttgg 600

ccacttttat gagcgctgca tccgggggggc cacggccgat gcagcccgta cccgggagtt 660

cctggaaggc ttcgtggatg acttgctgga agccctgagg agcctctgca accgggacac 720

cgacatggag gtggaggact tcattggcgt ggacagcatg tacgagaact ggcaggtgga 780

caggccactg ctgtgccacc ttttcgtgcc cttcacaccc cccgagccct accgcttcca 840

cccagagctc tggtgctccg gccgctcagt gcccctggat cgccagggct acggccagat 900

caaggtggtc cgcgccgatg gggacacatt gagctgcatc tgcggcaaga ccaagctcgg 960

ggaagacatg ctgtgtctcc tgcacggcag gaacagcatg gcgcctccct gcggcgacat 1020

ggagaacctg ctgtgtgcca cagattccct gtacctggac acgatgcagg tcatgaagtg 1080

gttccagacg gccctcacca gagcctggaa gggcatcgcc cacaagtacg agttcgacct 1140

ggcctttggc cagctggaca gcccggggtc cctgaagatc aagttccgtt cagggaagtt 1200

catgcccttc aacctgattc ctgtgatcca gtgtgatgac tcggacctgt actttgtctc 1260

ccaccttccc agggagccct ctgagggcac cccagcctcc agcacagact ggctcctgtc 1320

ctttgctgtc tatgagcgac acttcctcag gacgacacta aaggcactgc ccgagggcgc 1380

ctgccacctc agctgcctgc agatagcatc cttcctgctc tccaagcaga gccgcctgac 1440

cggtcccagc gggctcagca gctaccacct gaagacggcc ctactgcacc tcctactcct 1500

ccggcaggcc gccgactgga aggcggggca gctggacgct cgtctgcacg agttgctgtg 1560

cttcctggag aagagcttgc tccagaagaa gctccaccac ttcttcatcg gcaaccgcaa 1620

ggtgcctgag gccatgggac tccctgaggc cgtgctcagg gccgagcccc tcaacctctt 1680

ccggcccttc gtcctgcagc gaagccttta ccgtaagaca ctggactcct tctatgagat 1740

gctcaagaat gccccagcgc tcattagcga gtattcccta catgtccccc cagaccagcc 1800

taccccaaaa agctgacgtc ttttacagaa tgtgggatcc tcgagctaag atgagggcat 1860

ccctcacgtt cacacccctg gtggcatctg ccagcgctgt tctggggaca aggcgggctt 1920

tcgtgggagc cgtgctcagc ctgccaggaa gccaagccct acagtgcaga ggaaacagaa 1980

tttcaacggg aagctggttt gcttcatacc attgggatct gctggtaaag ctgttatttg 2040

ggtttaggga ctgatccctt gcagtttact tctggatcac catgaatggc caagatggtg 2100

gcagaacacg ctgtggaccc tgagttagag acaatgcaaa tgttggattg ggtgtaattc 2160

tttttgaatc ccagatccag tctgtacttg aatatgagca gaggatctac aagaatgctg 2220

acagggaacc gtgttaagac ccagcacccc tattcccagg agcttctggc ctgaccatct 2280

gcagccaaag cactaacagg gacagatatg ggaatgtcca cctttgatcc gcatcctgca 2340

caatagtggt cccaccatgg ctgccacttt tttatactat ttggagaaaa gaccttgtat 2400

aaattcgagg cccgagtgac taacgtctct gtcacacgga aatgggtact tggtggcata 2460

gagaaacaca attagccact ttttcagcta cacttctcac tcagctgcac cctacacttc 2520

tcactcaggt gcacccccct ctgctgtcct ttccccaacg tactgggtcc cgagcgtggt 2580

gggtatttgc cacactgggt gccagctcag cagcccccca cctctcttta ttctctccaa 2640

agctggtctt tctgactatc attgtggtag ggggaggaca gatgctaaag gtggaagctg 2700

acctggaaaa agagacacac ggggtgactg tggcaaagga cagctggaaa agaaactcta 2760

tcacttcttc attggcaacc acaaggcacc tgaggccatg gcactcccag aggctgtgcg 2820

cagagccaag cctctcaacc tcttctggcc ctgcgtcctg cagcgaagtc tctgctgtaa 2880

gacagtagac tccttcgatg aggtgctcaa aaatgctacc cggggtggtg gtgtctggct 2940

tgcagtctgg cccagttcag agaaagttgc agagatcagg ggccaaggat gtcatagccc 3000

caggttgtcc tcagggtccc aatcctaggg cagggtgtgc atggaagcaa gaactatgga 3060

aacctagctc cagtctgcag gctctgagcc cctagttcct cactccagcg gggctccctc 3120

actgcacaga acccacccct tctgtgtggg cactgctgac cacacagatg acccagaccc 3180

aaagagcctg gcagaagctc tgtggttgga gctgggctcc gtctccaggt ctggttcagg 3240

gggatcagga aggctctttt ccacctgtgg cttcactggc cctttgagat ttcctatctc 3300

accgttactt cagttaccct tgcaggggge cagggagtca agaatatacc gtgttcctcc 3360

agggtttaag ccggccatgc cttcccgaga gcataaccaa cttgacaggg gtgcccagtt 3420

accccacaaa ctgaaggaag gagatccttc ccccatcccc aggagtgctc tcaaccagcc 3480

tcagaaagct tgagaagatg gacccttttgc ccaccagggt taattcctgg tggggcagct 3540

cggctgtgat cagggcaacc aaacctatag gaagccttcc agtgtgagct ggaattagac 3600

tgaacatgtg cttgggcctg cctctcccta gacgcagttg cggggcactc cagggaatga 3660

accagctcaa gtgtgtccct aacagcagcc tggagctacc cccaatccct cacagcctga 3720

ccctcctcat tccatcagat gcatttgtag aatcggggca aatttcctta tttatttatg 3780

gcccatgcct ttccccctct tcatcctgat cccgttttgc tttgaagaga ccccagtaac 3840

caaaaaacag cctccagaag ccaaaaccat gcctggatct cccatagctt ctcctttgct 3900

tccaggagaa agttcactga aaaaaaaata tcttctggct tcttgtgtgt acagagacaa 3960

cagaactcgg tggggaaacg ggaatctttt ctgcaccaaa gctgcttcta aagcagaaag 4020

cagtggggct cttggtgttt catgctgcct tatttatatt aaaggaagaa ttaaatcttg 4080

caaggagtaa aaatggtcac tgtttggttt ttaaacttca ggaaatctgt gaggctccca 4140

gggcagagtt ggggacaggg gggtggattt cctattgaca aagcagaagc tttcactccc 4200

tttttattct tacactttca ggagacttta aaaaataaca aaacagaaca tttctaccct 4260

gtctttgagg caacttgtgt gctgcccgca gggctccaga aagggccttt gaaagctggc 4320

cgtagggtgg attccagcct aagaccttat ttcacagagc taatgtcagc gagagatata 4380

ggagctggaa ctaacaggag cgtttggttt tagagaaaga ggaggtgtct tggataactc 4440

t 4441

<210> 116

<211> 120
<212> DNA
<213> Homo sapien

<400> 116
acttacaaac tgtggtcatg ttaagataca gaacatcagc caagaaccag aaccaactcc 60

ttaacataac atagagtggg ccctgcccac cccccaatac aatacattta agataagctt 120

<210> 117
<211> 1977
<212> DNA
<213> Homo sapien

<400> 117
ttcacccggc cctcgcctgt caccttcaca gagaaggggc tgcctgcagg aagaaagcac 60

ggcccacgcc cctccagtca catactgcct gtgggccctg gtgtagtgag gggggctgcc 120

gaggcactgc tgcactcacc aggcacgtgc tggtcggcaa acttgatgtt gatgatgtag 180

ttgcctggct ctgtggggca gtaggtcacc tggcgggaca ggtgcctggc tgctacaaac 240

catgcaatga gccatgcccc gccctggaca ccccgccca gcatctgggc ctccacgctt 300

gggaccgtgg gagcggccaa cagagctatg tctggagaca tatgataaac cacctcagcc 360

cccaccaagc cgccgcaccc gtagaccaga ccccaaggac cctggccacc atgggccaga 420

gagcattacc ttcatctctg gctctgctga gccggccctt gagtccccca cctgctgcct 480

gctctggcga ccctgggtgt gggagtggtg ccgggctgcc ttctgcttcc gcccgctgcc 540

gggattgcct ccagcgctgt ggaggcctgt gtgcggggat gcagcccctg cctgtctact 600

gaggactccc actgagggga ctgctgaagc caactggtgc caaggagcac aatggagtgc 660

cccccagccc tgatcgtgca ccccccagac cggcgggatg gccaggcggg ctgcaagtca 720

accatgggca gcagcttcag ctaccccgat gttaagctca aaggcatccc tgtgtatccc 780

tacccgagag gccacctccc cagcccctga tgcggactcc tgctgcaagg agccactggc 840

cgatccccca cccatgcgag cacagcctgc ccagcacctt tgccagtagt cctcgtggct 900

ccgaggagta ctattctttc catgagtcgg acctggacct gccggagatg ggcagtggct 960

ccatgtcgag ccgagaaatt gatgtgctca tcttcaagaa gctgacagag gctgttcagc 1020

gtacaccaga tcgatgagct ggccaagtgc acatcagaca ctgtgttcct ggagaagacc 1080

agtaagatct cggaccttat cagcagcatc acgcaggact accacctgga tgagcaggat 1140

gctgagggcc gcctggtacg cggcatcatt cgcattagta cccgaaagag ccgtgctcgc 1200

ccacagacct cggagggtcg ttcaactcgg gctgctgccc caaccgctgc tgcccctgac 1260

agtggccatg agaccatggt gggctcaggt ctcagccagg atgagctgac agtgcagatc 1320

tcccaggaga cgactgcaga tgccatcgcc cggaagctga ggccttatgg agctccaggg 1380

tacccagcaa gccatgactc atccttccag ggcaccgaca cagactcgtc gggggcaccc 1440

ttgctccagg tgtactgcta acccctgcca ggcccagctg ccacaccctt tctgggagaa 1500

gcatggccta cagaatgaag agggggacca ggaacccctg tgggagaggc ttagacctga 1560

agcagtgccc actctggctc ctcctgcctt ggctgactgg gttcctggac catgtgcatt 1620

tcactgggcc atgggatcta catctccttg catccccagc tggtctgatc cctgccaggg 1680

ccccttcctt cctgctcatg gtcttcaggt ggcctgatca tggaaagtaa ggagttaggc 1740

attaccttct gggagtgaac cctgactcca tccccctatt gccaccctaa ccaatcatgc 1800

aaacttctcc ctccctgggg taattcaaca gttaaaagaa gcttatctta aatgtattgt 1860

attggggggt gggcagggcc cactctatgt tatgttaagg agttggttct ggttcttggc 1920

tgatgttctg tatcttaaca tgaccacagt ttgtaagtac ctcgctcgcg accacgc 1977

<210> 118
<211> 182
<212> DNA
<213> Homo sapien

<400> 118
catgttctct aatgcatgct ccgacggcgc atgtgatgga tcgcggcgag tgaaaagcaa 60

gagccagaat taagaggttg ggtcagtctg gcagtgagtt catgcattta gaggtgttct 120

tcaagatgac taatgttcaa aaattgagac atctgttgcg gttcttttt ttttttttc 180

cc 182

<210> 119
<211> 875
<212> DNA
<213> Homo sapien

<400> 119
ggtcgcggcc gaggtaccac attggtccac ttgacactaa ccaatcgatc attttttttt 60

aatcaagaaa gctagattct atcagataaa atcactgctt ctaaagagtt taaatctagt 120

tagaaaaagt tatagaaatg tttgcaaaga taagtaacag atagagtcag tagaggataa 180

gatcaaaaac aaaaccaagc aaaagatgag ttcaggggag tttgccatca agtggcaaaa 240

ctgacttact tagggaagaa agttataaaa caggaaaata tgagatgaac cttgagtgat 300

gtggaagatt tagataaatg gaaaggaagg agaaaatgga gttctttagg tggttgtaat 360

tggaggagga aatgaataca cacatcttgt tgacttaaac ccagacattc agcagctctc 420

tatacatatc tggaaaagac tgcacagtca cctcctgtct ctcaccccag gtattactta 480

gaattattat catatttccc ttcctttaaa gtaagtaagg gtgatggtga caatatggag 540

aactatgatt tttccattaa cctaataata attgtattta ttgagttctg ttaagcattt 600

tacatattaa ctcacttaag cctttcaaca gcctgggcaa aataggtatt attatccccc 660

attttacagg caaagaaaac tgaggtttaa ggtaactgtg ccgaaagtgc catataacag 720

ggctcacatt cagtatctgc agttgcaagc tcatgatcta tagtgccaag ttgcatatgg 780

tagtccatgt cacattatta ccctttttata tccctggaat tttcatgggc aaccattagt 840

attcatttta atatcactaa acttccagcc ctgat 875

<210> 120
<211> 987
<212> DNA
<213> Homo sapien

<400> 120
ggtcgcggcc gaggtaccac attggtccac ttgacactaa ccaatcgatc attttttttt 60

aatcaagaaa gctagattct atcagataaa atcactgctt ctaaagagtt taaatctagt 120

tagaaaaagt tatagaaatg tttgcaaaga taagtaacag atagagtcag tagaggataa 180

gatcaaaaac aaaaccaagc aaaagatgag ttcaggggag tttgccatca agtggcaaaa 240

ctgacttact tagggaagaa agttataaaa caggaaaata tgagatgaac cttgagtgat 300

gtggaagatt tagataaatg gaaaggaagg agaaaatgga gttctttagg tggttgtaat 360

tggaggagga aatgaataca cacatcttgt tgacttaaac ccagacattc agcagctctc 420

tatacatatc tggaaaagac tgcacagtca cctcctgtct ctcaccccag gtattactta 480

gaattattat catatttccc ttcctttaaa gtaagtaagg gtgatggtga caatatggag 540

aactatgatt tttccattaa cctaataata attgtattta ttgagttctg ttaagcattt 600

tacatattaa ctcacttaag cctttcaaca gcctgggcaa aataggtatt attatccccc 660

attttacagg caaagaaaac tgaggtttaa ggtaactgtg ccgaaagtgc catataacag 720

ggctcacatt cagtatctgc agttgcaagc tcatgatcta tagtgccaag ttgcatatgg 780

tagtccatgt cacattatta ccccttttta tatccctgga atttttccat gggcaaccat 840

tagctatttc atttaataat cacctaaaac ttttcagtct tctgattaaa attacgctgg 900

agtgatagaa tgtattttca tgatagaaat tgggaaaaaa aatggggaat gaagtttatc 960

agcatttcag acttgttttt ttttttt 987

<210> 121

<211> 295
<212> DNA
<213> Homo sapien

<400> 121
cgtggtcgcg gcgaggtaca taaagctatg attttgcatg ggaactcatt taaattaaag 60

ttaccaaggg aattccatat gtaaactttt cctggaacta tgaacttctg agacttcaga 120

aagattttga ctgtgtaggt tattctgtgt tgcgctatgg ggtccgcctt tgtgttgctg 180

agctggagag cgtgcctctg ctgccgcgct gtgtcagtag tggggattgc acttttgttt 240

ccagctacag gccaaatttg agcaccttga gagacttcac caagaagaga gaatg 295

<210> 122
<211> 3210
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (954)..(1013)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (1958)..(2054)
<223> a, c, g or t

<400> 122
tagtgtaaat tctactatga tgagtatttt tatatgtaaa gctttctctg aatttaggat 60

ttttttaagc ctagatctca ggacattgaa agactatgga aaaatagaat acagttttaa 120

cattcttgac atagttgaca attgacactt gctttccagt gatgtttact agtacagttt 180

tcaataactg tttactgcat atattttatt tggaaaatag tgttgttgtt tttttccttt 240

ctggtaattt ctttctatac tttgcatctc ttatttttat tgagcatgga cttgagagac 300

tttaagaagt tttgtaggga aaatacacac cgagaaaata aaaactctag gaaaagttca 360

caatgtttta aaaatagttt tgtacataaa gctatgattt ttgcatggga actcatttaa 420

attaaagtta ccaagggaat tccatatgta aacttttcct tgaactatga acttctgaga 480

cttcagaaag atttgactgt gtaggttatt ctgtgttgcg ctatggggtc cgcctttgtg 540

ttgctgagct ggagagcgtg cctctgctgc cgcgctgtgt cagtagtggg gattgcactt 600

ttgtttccag ctacaggcca aatttgagca ccttaagaga cttcaccaag aagagagaat 660

gaagcttgaa gaaaggagag acttttggaa gaagaaataa ttgctttctc taaaaagaaa 720

gctacctccg agatatttca cagccagtcc tttctggcaa caggcagcaa cctcaggaag 780

gacaaggacc gtaagaggta agaggcccag ccaccctcag gttgctgccc tgcccacact 840

tgtcacattg ccctgtcgga tcttcctagc gcttcttact tctgcttcct tatgctttgt 900

cttccttctc caactaaatg tcagctcctg gagaggagag accttgtccc tctnnnnnnn 960

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnntttgcca 1020

aaatgaatga atgaacaaaa taaaaataca tcacagtgtg aaaaggatca gttagctggg 1080

ggaaaatggc tgctcttatt gaatagccca gctgtggccc tagtacacag ctaatggctg 1140

ctgacctacc atgagttatt ttgaatttca tgtctaaata aagctgtgcc ctttgttggg 1200

gaattataga cataagatag tattttattg aattaattct taggttcagt tttttagaac 1260

aaattctact gatactgatt ttgtgagaat ttttatttta ataagggaac caggctgtcg 1320

attcgaactc ctgtgcattg atgtcagagc ttgtgaaacc aatggtggac gtaaagacgc 1380

agagaaaggt agagagggcc acagtgcaag ggggcagcca tgagtagcat gtggatgata 1440

atcgcaaggc gtgtaggttt ggggtatgtg ttaaaatcag ttacccaagt tttagaaaat 1500

aagtttattt ttccttttttg ataatttcct tctgtacttt gtgtcccttg ttttatttct 1560

cccacttaag tatccctgtc ggtttaacta tttggctctt gaggaaattc atcatgcctt 1620

gttctattca gaatgctcat ttgggggact ctgcctcagt ttactctagc tacctctgtg 1680

ctagcagcgt tgcatgctcc catgttcccc acagagcaac acatctggtt taatgctctc 1740

actgcatcta tttcagaccc attttcagaa tattgcaact caactttaaa ataggaaaac 1800

ataaccttgg cccattggtg gctttcaatt tttaaatatg gagaagtaaa ccacaggact 1860

ttggcagatt ttattggctt atatatttta caggtatttt tgtttgtttg ttatgttttt 1920

ttggaaaaga ggaaggaatt ctcatgtatg atctttcnnn nnnnnnnnnn nnnnnnnnnn 1980

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2040

nnnnnnnnnn nnnnttattt ttgaaagaat aaatagtata aaagctgtat tttattcaag 2100

cattgaattt agaaagataa actataaatt tattactgct tttcaacatg ctttggattt 2160

taaagctaat ggatctttta ttaatacctt ttttcctttc atcttaaccc agtgttacta 2220

aaattagatt tcccattttt tcccctatat gaaagataat ttacatttac cttgtaaaaa 2280

ttatcaccct gctccacttg agaaccctgc tgttgtttgc aaaatcagga ccaagcccat 2340

atttttgag agctctctag aagaatttat cttcaagaaa atatggtttg gtgtttttgg 2400

ttttttcttt tctttttcaa agtcccatgg ccagatatcc tataatatac tagatgcatg 2460

tttgctaatt tttacttgaa tttttttaa attgtaccaa tcaaaaggtt ctttttcttt 2520

ccagctccaa ttttttgtaa aacagaagtt ccagagcaca gaaggtcatc atcacaagca 2580

aactttatta aaaaaaaact agaagtgtgc tttgattttg ctgttatttg ttttatcact 2640

tctatatttg gtgaacagcc acagttactg atatttatgg aaaagtactt tcaagtacaa 2700

ggtcaataca taagccagag tgaatgatac tacaagttga gcatctctaa ttcaaaaatc 2760

tgaaatccag aagcttcaaa atctgaatct ttttgagcac tgacttgacc ccacaagtgg 2820

aaaattcccc acccgacacc tttgctttct gatggttcag tttaaacaga ttttgtttct 2880

tgcacaaaat ttttgtataa attactttca ggctatatgt ataaggtgga tgtgaaacat 2940

gaattatgta attagagtcg ggtcccgtag tgtatatgca gatattccaa acctgaaatc 3000

caaaacactt ctggtcccta agcattttgc actaagggat actcagcttg tacctatata 3060

ttcctatata ttcactgttg ttagaaatgt ttaagttgct gttctgtgat gaatctaaat 3120

cttttctctt gctaccaagc tattgtcact gcagtgcatt ataccaaaga gcgaagtcag 3180

tgccactgaa aatacagaac cattatatcg 3210

<210> 123
<211> 662
<212> DNA
<213> Homo sapien

<400> 123
gcgtggtcgc ggcgaggtac ctgcatctaa aatagaattt atttttcttg agacacgtct 60

ttcaggcaga ggaaaggtaa tgggcacaca aaccttgatc atcttcttat aaaacaaggc 120

agcatccaca tttcaatact gtagcatatg aaataatttc tcaattccta ctctgtcata 180

gcttaaatga cttactgatt aaaatatgag taacaataat tctggagctg aaagacattc 240

tcaagccaaa aggtaacatt aaacaatgtt ttactttaaa tttaatgtca tgtagggtca 300

ctaaaaacaa aatgtaaggg ataaatgtaa gtgcataaaa cagagataat accaatttcc 360

acaaatacaa cttttattat ccaaaatcat cttggaagga ctttttctaa tatgcccatt 420

ttctaaataa gattaaccat ttgatgggaa tatttccaat tgtatcacct ccttccttga 480

ctttttcttc atcaacttag ggcctgggtt gtaggcactg tggctctttg ggatagttaa 540

atgggtgcaa tttggctgag accgcccctg tcttaaatgg cccaggctac agggcttgcc 600

aaatggcttg tggtttgcaa cctcttcctt tgatatccat tgaagaacac atgtcctacc 660

ct 662

<210> 124
<211> 1845
<212> DNA

<213> Homo sapien

<400> 124

tgtcagattg atggcctgtg ctgttgagta gttccgtctg attgacgtgg cgattcggat 60

gtggtcgggt gtggtgagtg gctgggtggg ctttcggtgg gtggtgtgtg tgcggccggt 120

ggtttgtgtg tgcgtgtgtc gcggcggtgt gggctgtggg tcggacggag ggcctggtgg 180

cgttgttaag aggctcgggt gtcggcgggg gtgtgttctg ttgggggctc gtcgtgtccg 240

ttcggtggtc cggcgtggtt gtgtggatat cgtggtgtcg cgtggttgcg tggctgcgct 300

ggtgcgggcg tggcttatgt tgtgttttaa tgcgttgctc gtggggtcgg tggggtcgtg 360

gcggtgtggt ggtgtgtatg tgtggtgttt gttcgcatgg ggcgtgttgt atgtgttggt 420

ggccgcggta ggagtccggc catctcctga tttattcttt tacagttttc tatgttttga 480

attgatatca attcttagaa taaaatgcag tacctgcatc taaaatagaa tttatttttc 540

ttgagacacg tctttcaggc agaggaaagg taatgggcac acaaaccttg atcatcttct 600

tataaaacaa ggcagcatcc acatttcaat actgtagcat atgaaataat ttctcaattc 660

ctactctgtc atagcttaaa tgacttactg attaaaatat gagtaacaat aattctggag 720

ctgaaagaca ttctcaagcc aaaaggtaac attaaacaat gttttacttt aaatttaatg 780

tcatgtaggg tcactaaaaa caaaatgtaa gggataaatg taagtgcata aaacagagat 840

aataccaatt tccacaaata caacttttat tatccaaaat catcttggaa ggacttttc 900

taatatgccc attttctaaa taagattaac catttgatgg gaatatttcc aattgtatca 960

cctccttcct tgacttttc ttcatcaact tagggcctgg gttgtaggca ctgtggctct 1020

ttgggatagt taaatgggtg caatttggct gagaccgccc ctgtcttaaa tggcccaggc 1080

tacagggctt gccaaatggc tttggtttgc aacctcttcc tttgatatcc attgaagaac 1140

acatgtccta ccctcttgat tgcaagcttc tatctgctat tggctatcag gagcataaca 1200

gataagttct aaggtccttc cagttccaaa agccattgac caaccctgtg aggcactatg 1260

agagattgga gattactgag gaatctcttc caataacata tgatatataa aggtaattgt 1320

gattctatga agagtgaata aaaattgaaa acaaaaccaa tcttgtttaa actattaaat 1380

gaagttttaa aaaataatgg agtgggtaga tgtattgcct atcaaattaa aattcaggga 1440

acacgtgtca gttatcacct ttataagtat gatctataag gttgaattaa ctaaagatct 1500

aaccctaaac taagtgagat attactgggt tactaaatat agatgaatgg cttccttttt 1560

cctttaaaga tggaagaatc tcaaaagtat gcctgaatca gaagcagagg actcatgttc 1620

taatgacacg tctgatcatc accctgagtt ccataaagct atagcatatt gtagtagaaa 1680

gtatatcagc atgtgagtgg agagatttaa gggtatagat ctctacgtat ttccagctct 1740

gtcaccaact ggtgacatga ctttgggcag aaactgtctt attgattccc tcgtgccgaa 1800

ttcttgctcg aggccaaatt ccctatagcg aacttattga acaga 1845

<210> 125
<211> 306
<212> DNA
<213> Homo sapien

<400> 125
gcggcgccgg gcaggtacat aagtgtcact cgcctaatta aaaacattga gtaaaccaag 60

tttttatata gactaccctt gccatatgat gctctttttc tctaataata tgcagtttaa 120

atcctgagga atcaatgccc agcatttcac cacatctgaa ctctgtgtgg gcattcttca 180

ctcgcctaca aggggtaaac aaggctacca gaacttgaat ttgacttata gggagctacc 240

caggaagggg aaagcccttg ggacttttc caaaacaatc ttctatttga actgttcatc 300

agccaa 306

<210> 126
<211> 2049
<212> DNA
<213> Homo sapien

<400> 126
aaaaaaaaaa gtaaaacaaa aataaagtct atgcccatta agacgtcttc taattcagtt 60

gtgattgtct gctcctactt aaaaaaatat ttaagcttga tgtttaatta ttccctttca 120

gcaaatttgg atcagaaaat taaagtatgt gacaagatca ggtcaccttg aatttccaca 180

caatctcaag acactgaata gcaaaaaagt aacattacac agtaatgatt aggatatttc 240

cttagacttt gctggatctt tggtcttaag gtaacatgta aaagtagtga agcctttcct 300

ttcatggccc tgtgcaatgt aacggttttc tgcctcctct tcagctggaa gcgttagtgg 360

tagtatgggc acagaatata tgtacactgg cgatgctgac catgcctccc aggtaccctg 420

gctctgggtt ccttgaccta gggaacaaga ttggatgagg cagatctttg agcccatgtg 480

actatagaat ttgctgatga tataatttta caataacaat ggataggaat tttacctctc 540

tttttattag tttaatatta tttaatatta tgtacataag tgttcactcg cctaattaaa 600

aacattgagt aaaccaagtt tttatataga ctacccttgc catatgatgc tctttttctc 660

taataatatg cagtttaaat cctgaggaat caatgcccag catttcacca catctgaact 720

ctgtgtgggc attcttcact cgcctacaag gggtaaacaa ggctaccaga acttgaattt 780

gacttatagg gagctaccca ggaaggggaa agcccttggg acttttttcca aaacaatctt 840

ctatttgaac tgttcatcag ccaaagtagt ccactgaggt gacaaagctt tcagaaatac 900

aaagatggga agataaaggt aacactggcc cacttggggc tttgacattg gattgggtgg 960

actgaataaa cacagcctag gtggcctggg cttgagcctc acttacttct ccttgataca 1020

tagttcctgg tctaccttct gacccttttt ctaaaatagc cagtgtctat ttcactaggc 1080

catttactta caagttccca gcttttaggg aaaaaagagg gaggggggag catctagttt 1140

tgaattagat atacatctta gaagtaatga gctattggca gctgttaaat cagattcagc 1200

cacaaaccag aattctttct tgttgaacaa gaccaatgag ttagatgact ttaataattc 1260

cacttttctc tccctctctt ctcctcttcc tgaaatcaga gagatgagaa actactcttt 1320

gaaatacctc cagaggcgtt ttattgtgtt cctttccccc ccaagcagct ccctttatac 1380

aattttgctc aggcaaccaa ggacagagta tcggcagaaa catggagtgc ttttgtatag 1440

gccacctgta cataaaagtg taattattta tttaattttc ccatttgtat catattaaag 1500

ctttgtacag tgttttaagt tctgttttaa aattattttg tattttattt ttataaccta 1560

gtaataaaat attcattccg catgcaaaaa aaaacacaca cacaacccaa acaaccaaaa 1620

acaaacagaa ccaaaagata gaaaccaaac acagcaacag acagacacaa gagaagacga 1680

aaacaaccac acaaacagca cacacaaccg cagcggagaa ccaacaaagc caaacgcaag 1740

acagcacaac aagggacaaa acacacacgt agcacaaagc agccgcagaa ccgaacacac 1800

atctaaagac aggcgacaac aacatagcag gccgtagcac cgacaaccac taaaccataa 1860

ctatccagcg gagatcagat cacagaagcc gacacagaaa gaaaagcatg tatgcaacgc 1920

atgacaccgc agggcggata aaaacactcg taggcgggac cgccgcagaa cgatcgaagg 1980

ctaacaaaca agctagtaga tcacaagtac gtgaccacgc cctcttatct ataaaggagg 2040

aagagggag 2049

<210> 127
<211> 286
<212> DNA
<213> Homo sapien

<400> 127
acatctggtt gcggagaaag gaaaaacttc tcaaagaaga atttgaagtg agctttaagg 60

gttgaatgga atttttactg ttagaggtag agaaatacaa tatcataaag aaagatgtga 120

tccctacaag aggattacgt ggtaagttga aagacattaa acagtccaac ctggtcattg 180

tcaaaactat atacgttggg cacagaactg aagaccaggt gtcaaaggag gatggctctg 240

tccccttttgt gtccccagtg cctaaagcag tgtttggtgc aagctt 286

<210> 128
<211> 12421
<212> DNA
<213> Homo sapien

<400> 128
agcccacggg atcgcccacg cgtcggggca ggtcaaggag gaggctgaaa gagcctgagc 60

tgtgccctct ccattccact gctgtggcag ggtcagaaat cttggataga gaaaaccttt 120

tgcaaacggg aatgtatctt tgtaattcct agcacgaaag actctaacag gtgttgctgt 180

ggccagttca ccaaccagca tatcccccct ctgccaagtg caacacccag caaaaatgaa 240

gaggaaaaca aacaggtgga gactcagcct gagaaatggt ctgttgccaa gcacaccaca 300

gagctaccca acagattcct atggagttct tgaattccag ggtggcggat attccaataa 360

agccatgtat atccgtgtat cctatgacac caagccagac tcactgctcc atctcatggt 420

gaaagattgg cagctggaac tccccaagct cttaatatct gtgcatggag gcctccagaa 480

ctttgagatg cagcccaagc tgaaacaagt ctttgggaaa ggcctgatca aggctgctat 540

gaccaccggg gcctggatct tcaccggggg tgtcagcaca ggtgttatca gccacgtagg 600

ggatgccttg aaagaccact cctccaagtc cagaggccgg gtttgtgcta taggaattgc 660

tccatggggc atcgtggaga ataaggaaga cctggttgga aaggatgtaa caagagtgta 720

ccagaccatg tccaaccctc taagtaagct ctctgtgctc aacaactccc acacccactt 780

catcctggct gacaatggca ccctgggcaa gtatggcgcc gaggtgaagc tgcgaaggct 840

gctggaaaag cacatctccc tccagaagat caacacaaga ctggggcagg gcgtgcccct 900

cgtgggtctc gtggtggagg ggggccctaa cgtggtgtcc atcgtcttgg aatacctgca 960

agaagagcct cccatccctg tggtgatttg tgatggcagc ggacgtgcct cggacatcct 1020

gtcctttgcg cacaagtact gtgaagaagg cggaataata aatgagtccc tcagggagca 1080

gcttctagtt accattcaga aaacatttaa ttataataag gcacaatcac atcagctgtt 1140

tgcaattata atggagtgca tgaagaagaa agaactcgtc actgtgttca gaatgggttc 1200

tgagggccag caggacatcg agatggcaat tttaactgcc ctgctgaaag gaacaaacgt 1260

atctgctcca gatcagctga gcttggcact ggcttggaac cgcgtggaca tagcacgaag 1320

ccagatcttt gtctttgggc cccactggac gcccctggga agcctggcac ccccgacgga 1380

cagcaaagcc acggagaagg agaagaagcc acccatggcc accaccaagg gaggaagagg 1440

aaaagggaaa ggcaagaaga aagggaaagt gaaagaggaa gtggaggaag aaactgaccc 1500

ccggaagata gagctgctga actgggtgaa tgctttggag caagcgatgc tagatgcttt 1560

agtcttagat cgtgtcgact ttgtgaagct cctgattgaa aacggagtga acatgcaaca 1620

ctttctgacc attccgaggc tggaggagct ctataacaca agactgggtc caccaaacac 1680

acttcatctg ctggtgaggg atgtgaaaaa gagcaacctt ccgcctgatt accacatcag 1740

cctcatagac atcgggctcg tgctggagta cctcatggga ggagcctacc gctgcaacta 1800

cactcggaaa aactttcgga ccctttacaa caacttgttt ggaccaaaga ggcctaaagc 1860

tcttaaactt ctgggaatgg aagatgatga gcctccagct aaagggaaga aaaaaaaaaa 1920

aaagaaaaag gaggaagaga tcgacattga tgtggacgac cctgccgtga gtcggttcca 1980

gtatcccttc cacgagctga tggtgtgggc agtgctgatg aaacgccaga aaatggcagt 2040

gttcctctgg cagcgagggg aagagagcat ggccaaggcc ctggtggcct gcaagctcta 2100

caaggccatg gcccacgagt cctccgagag tgatctggtg gatgacatct cccaggactt 2160

ggataacaat tccaaagact tcggccagct tgctttggag ttattagacc agtcctataa 2220

gcatgacgag cagatcgcta tgaaactcct gacctacgag ctgaaaaact ggagcaactc 2280

gacctgcctc aaactggccg tggcagccaa acaccgggac ttcattgctc acacctgcag 2340

ccagatgctg ctgaccgata tgtggatggg aagactgcgg atgcggaaga accccggcct 2400

gaaggttatc atggggattc ttctaccccc caccatcttg tttttggaat ttcgcacata 2460

tgatgatttc tcgtatcaaa catccaagga aaacgaggat ggcaaagaaa aagaagagga 2520

aaatacggat gcaaatgcag atgctggctc aagaaagggg gatgaggaga acgagcataa 2580

aaaacagaga agtattccca tcggaacaaa gatctgtgaa ttctataacg cgcccattgt 2640

caagttctgg ttttacacaa tatcatactt gggctacctg ctgctgttta actacgtcat 2700

cctggtgcgg atggatggct ggccgtccct ccaggagtgg atcgtcatct cctacatcgt 2760

gagcctggcg ttagagaaga tacgagagat cctcatgtca gaaccaggca aactcagcca 2820

gaaaatcaaa gtttggcttc aggagtactg gaacatcaca gatctcgtgg ccatttccac 2880

attcatgatt ggagcaattc ttcgcctaca gaaccagccc tacatgggct atggccgggt 2940

gatctactgt gtggatatca tcttctggta catccgtgtc ctggacatct ttggtgtcaa 3000

caagtatctg gggccatacg tgatgatgat tggaaagatg atgatcgaca tgctgtactt 3060

tgtggtcatc atgctggtcg tgctcatgag tttcggagta gcccgtcaag ccattctgca 3120

tccagaggag aagccctctt ggaaactggc ccgaaacatc ttctacatgc cctactggat 3180

gatctatgga gaggtgtttg cagaccagat agacctctac gccatggaaa ttaatcctcc 3240

ttgtggtgag aacctatatg atgaggaggg caagcggctt cctccctgta tccccggcgc 3300

ctggctcact ccagcactca tggcgtgcta tctactggtc gccaacatcc tgctggtgaa 3360

cctgctgatt gctgtgttca acaatacctt ctttgaagta aaatcaatat ccaaccaggt 3420

gtggaagttc cagcgatatc agctgattat gacatttcat gacaggccag tcctgccccc 3480

accgatgatc attttaagcc acatctacat catcattatg cgtctcagcg gccgctgcag 3540

gaaaaagaga gaaggggacc aagaggaacg ggatcgtgga ttgaagctct tccttagcga 3600

cgaggagcta aagaggctgc atgagttcga ggagcagtgc gtgcaggagc acttccggga 3660

gaaggaggat gagcagcagt cgtccagcga cgagcgcatc cgggtcactt ctgaaagagt 3720

tgaaaatatg tcaatgaggt tggaagaaat caatgaaaga gaaactttta tgaaaacttc 3780

cctgcagact gttgaccttc gacttgctca gctagaagaa ttatctaaca gaatggtgaa 3840

tgctcttgaa aatcttgcgg gaatcgacag gtctgacctg atccaggcac ggtcccgggc 3900

ttcttctgaa tgtgaggcaa cgtatcttct ccggcaaagc agcatcaata gcgctgatgg 3960

ctacagcttg tatcgatatc attttaacgg agaagagtta ttatttgagg atacatctct 4020

ctccacgtca ccagggacag gagtcaggaa aaaaacctgt tccttccgta taaaggaaga 4080

gaaggacgtg aaaacgcacc tagtcccaga atgtcagaac agtcttcacc tttcactggg 4140

cacaagcaca tcagcaaccc cagatggcag tcaccttgca gtagatgact taaagaacgc 4200

tgaagagtca aaattaggtc cagatattgg gatttcaaag gaagatgatg aaagacagac 4260

agactctaaa aaagaagaaa ctatttcccc aagtttaaat aaaacagatg tgatacatgg 4320

acaggacaaa tcagatgttc aaaacactca gctaacagtg gaaacgacaa atatagaagg 4380

cactatttcc tatcccctgg aagaaaccaa aattacacgc tatttccccg atgaaacgat 4440

caatgcttgt aaaacaatga agtccagaag cttcgtctat tcccggggaa gaaagctggt 4500

cggtggggtt aaccaggatg tagagtacag ttcaatcacg gaccagcaat tgacgacgga 4560

atggcaatgc caagttcaaa agatcacgcg ctctcatagc acagatattc cttacattgt 4620

gtcggaagct gcagtgcaag ctgagcaaaa agagcagttt gcagatatgc aagatgaaca 4680

ccatgtcgct gaagcaattc ctcgaatccc tcgcttgtcc ctaaccatta ctgacagaaa 4740

tgggatggaa aacttactgt ctgtgaagcc agatcaaact ttgggattcc catctctcag 4800

gtcaaaaagt ttacatggac atcctaggaa tgtgaaatcc attcagggaa agttagacag 4860

atctggacat gccagtagtg taagcagctt agtaattgtg tctggaatga cagcagaaga 4920

aaaaaaggtt aagaaagaga aagcttccac agaaactgaa tgctagtctg ttttgtttct 4980

ttaatttttt ttttttaaca gtcataacca ctaatgggtg tcatcttggc catctaaaca 5040

tcatcaattt ctaaaaacat tttccttaaa aaattttgga aattcagact tgatttacaa 5100

tttaatgcac taaaagtagt attttgttag catatgttag taggcttagt tttttccagt 5160

tgcagtagta tcaaatgaaa gtgatgatac tgtaacgaag ataaattggc taatcagtat 5220

acaagattat acaatctctt tattactgag ggccaccaaa tagcctagga agtgccctcg 5280

agcactgaag tcaccattag gtcactcaag aagtaagcaa ctagctgggc tcagtggctt 5340

atgcctgtaa tcctagcact ttgggaggct gaggtaggcg gattgcttga ggccaggagt 5400

tcaagaccaa cacggccaat gtggcgaaac cccgtctcta ctgaaaaata caaaattagc 5460

cgggtgtggt ggtgggtgct tgtagtctca gctacttggg aggctgaagc aagagaattg 5520

cttgaatcca ggaggaggag gttgcagtga gcccagatca tgccactaca ctccagcctg 5580

ggtgacagag tgaggcactg tctcaaaaaa aaaaaaaaaa aaaaaaaaga tactgaacta 5640

ggcacttgac atccaagatg tccacttggc ccttttctaa gtgtacttca ctttgttttc 5700

attcctcctc taaagctttt taacaaactg tcactcatgt tctaaaaaca aacaaatgtc 5760

tttcatctca ttttttactc agcattatct atcaaaggga ctggatgtcc catctagttt 5820

gtactttagt ttctagctca ttacatagga aatgcaaata aaaaacacaa tgagatacca 5880

cttcacaccc actaagatgg ccataactta tttttttttaa aaaaaggaaa aggaaaataa 5940

taagtgttgg tgaggaagtg gacagaaatt ggaattctca tacattgctg gtgggaatgt 6000

ggaatggttc aactgctgtg gaaaacagtt tgtctgttcc tcaaaaagct aaacatagaa 6060

ttaacatatg acccagcagt tccactccta gatatatacc caagagaatg caaaataggc 6120

attcaaacaa atacaggtac acaaacattc atagcagcac tattcccaat agccaaaagt 6180

cagaaacaac ccaaagtcca tcaatgaatg agtggataaa taaattgtgc tataaacaca 6240

taatacaatt ttatttggcc ataaaaatga agtgttgata catgctacta tgtaaataaa 6300

tctcaaaaac ttgtactaag tgaaagaagc cagacacaaa aggtctcata caattcgatt 6360

tatataaaat agcaagaata gataaaacca tagagacaga aagcagattg gtggttgctg 6420

ggggctggag gaaggaagga atggagagta tctgcttaat gggatggggt ttccttttag 6480

aggagaacat ttttttgga actttaaata aaggtggtgg ttatgcaaca ttgttaacaa 6540

tactaagtgg cactgaactt ctcactttaa aatagttaat tttatgttat gtgaagttta 6600

cctcaataaa aaaattctta aaaaaaaaaa ctcagtgata tccaaccatt aaaaaatata 6660

agttacatag ctcgtatata atcaacttac ttattactgc actatgaggg gcacccagat 6720

atgtaagaca gagtcccaac ctttggttgt tggaggacta tctaaggata cagagattaa 6780

taaaaataaa agaactacca cgaaagaagt gccctccaag tacaaagaga gttcaaatat 6840

gtacatctgg ttgcggagaa aggaaaaact tctcaaagaa gaatttgaag tgagctttaa 6900

gggttgaatg gaatttttac tgttagaggt agagaaatac aatatcataa agaaagatgt 6960

gatccctaca agaggattac gtggtaagtt gaaagacatt aaacagtcca acctggtcat 7020

tgtcaaaact atatacgttg ggcacagaac tgaagaccag gtgtcaaagg aggatggctc 7080

tgtccccttt gtgtccccag tgcctaaagc agtgtttggt gcatgcttgg tgattggtaa 7140

atgtcaaaga aaaagataaa agataaaata tagccgggcg cggtggctca agcctgtaat 7200

cccagcactt tgggaggccg agcttaacat tgagaaaaga aaggaaatgc aacaagaaaa 7260

gcagaaagca cttgatgtag aagcaagaaa gcaggttaac aggaagaaag ctttactgac 7320

tcgtgtccag gagattcttg acaatgttca ggttagaaaa gcacctaatg ccagtgattt 7380

tgatcagtgg gagatggaaa cagtttactc taattcagaa gtcagaaact tgaatgttcc 7440

tgctacattt ccaaatagct ttccaagcca tacggaacac tctactgcag caaagcttga 7500

taagatagct gggattttgc cattggataa tgaggaccaa tgtaaaactg atggaataga 7560

cttagctaga gattcagaag gatttaattc tccgaagcaa tgtgatagtt ccaatattag 7620

tcatgtagaa aatgaagctt ttccaaagac ctcttcagca accccacaag aaactcttat 7680

ttctgatggt cccttctcag taaatgaaca acaggatcta ccacttttgg cagaagtcat 7740

cccagatccc tatgtaatga gtcttcagaa tctgatgaaa aagtcaaagg aatatataga 7800

aagagaacaa tctagacgca gtctgagagg tagtatgaac agaattgtta atgagagtca 7860

tttagacaaa gaacatgatg ctgttgaagt ggctgactgt gtaaaagaga aaggccagtt 7920

gacaggcaaa cactgtgtct cagttattcc tgacaaacca agccttaata aatcaaatgt 7980

tcttctccaa ggtgcttcca ctcaagcaag cagcatgagt atgccagttt tagctagctt 8040

ttcgaaagtg gacataccta tacgaactgg ccatcccact gttctagagt ctaattctga 8100

ttttaaagtt attcccactt ttgttaccga aaataatgtt atcaaaagtc ttacaggttc 8160

atatgccaaa ttacctagtc cagagccaag tatgagtcct aaaatgcacc gaagacgttc 8220

caggacatca tcagcgtgtc acatacttat aaataaccca ataaatgcct gtgaattaag 8280

ccctaaagga aaagaacagg caatggactt aattattcaa gatactgatg aaaacacaaa 8340

tgtgcccgaa attatgccaa agttaccaac tgatttagcg ggagtttgtt caagcaaggt 8400

ttatgtgggc aaaaatacat ctgaagtcaa agaagatgtg gttttaggta aatcaaatca 8460

ggtatgtcaa tcttcaggaa atcatttaga aaataaagtt actcatggac ttgttactgt 8520

ggaaggtcag ttaacatccg atgagagagg cgcacacata atgaacagta cctgtgctgc 8580

gatgccaaag ctgcatgaac catatgccag cagtcagtgt atagcaagtc caaactttgg 8640

aactgtgagt ggactcaagc cagccagtat gttagagaaa aactgcagtt tgcaaacaga 8700

actgaataag tcttatgatg taaaaaaccc ttctccttta ttgatgcaaa accagaatac 8760

gagacagcag atggacacac ctatggtgtc ctgtggaaat gaacaatttt tggataacag 8820

ttttgagaaa gttaaacgga gacttgattt agatattgat ggtttgcaaa aagaaaactg 8880

cccttatgtc ataacaagtg gaataactga acaagaaagg caacatttgc cagaaaaaag 8940

ataccctaag ggatctggct tcgttaacaa gaataaaatg ttaggaacta gttccaaaga 9000

aagcgaggag ttactaaaaa gcaagatgtt agcttttgaa gaaatgcgga agagactaga 9060

agaacagcac gcccagcaat tatcactact catagctgag caggaaaggg aacaagaaag 9120

actgcaaaag gtttttagtc tggaaataca agcaaaattt aacaaaataa ctgcagtggc 9180

aaaaggattt cttactcgta gacttatgca gacagataag ctgaagcaac ttcgacaaac 9240

tgtaaaattg cgagctgcct tgtacggtat tcatgacata ttctttgtaa tggatgcagc 9300

tgaaagaatg tctattctac atcatgatcg agaagttcgc aaagagaaaa tgctcaggca 9360

aatggataaa atgaaaagtc cacgagtggc tctttcagct gcaacacaga agtctcttga 9420

taggaagaaa tacatgaatg ttgggaatag caggtacgag tggatatact ttcctactca 9480

tactgagctg aacactggag ggcccgttct ggagatgatc cttttccctg tcataggtca 9540

ccatctcatg aacaaccaca gaccaccacc cccgaatgaa aacatggcgt ctgcagaagg 9600

tgctgtgcgg ccgctgcggg gcggccagct gcccttcaca aacatggcgg ccgaggggtg 9660

cggggagtgg cggggtaagg atgggaagcc gagcagacgg ccccagaaca agcggtcatg 9720

tgactgggaa gatggccgtc tttccttgtt tattattaag gatattacaa agaatacaga 9780

tgaacagcca gatgaagagc tacgtagggt cacagagtca aagacaaaga aagtgaaccg 9840

gaaaggaagc acttcttcca cgtcctcctc ctcctccagc tccgtggtgg acccgctgag 9900

cagcgtcctc gatgggactg acccctctc catgtttgca gccactgctg accccgcagc 9960

cttggcagct gccatggaca gctccagaag gaaacgtgat agagatgata actccgttgt 10020

aggatcggat tttgagcctt ggaccaacaa acggggagaa atccttgccc ggtacaccac 10080

taccgaaaag ctgtctatta atctgtttat gggatctgaa aaaggcaaag ctgggactgc 10140

cacattggca atgtcagaga aggtgcggac ccggctggag gagctggatg actttgagga 10200

gggttcccaa aaggagctgt tgaacttgac tcagcaggat tacgtgaacc gcatagagga 10260

gctcaaccaa tcgctgaagg atgcctgggc ctcagaccag aaagtgaagg ctctaaaaat 10320

agtcatccag tgttcaaagc ttctttcaga caccagtgtt attcagttct acccaagcaa 10380

atttgtcctt atcaccgaca tacttgatac atttggaaag ctcgtgtacg agcgcatctt 10440

ttccatgtgt gtggatagcc gcagcgtctt accaggatgt ttttgtagtt acgtggaggc 10500

atccatcctg aaatgtaaca aattcctctc caaaacggga atttcagagt gcctgccccg 10560

gttgacatgc atgatcagag ggatcggaga cccactagtg tcggtgtatg cccgtgccta 10620

cctgtgccgg gtgggaatgg aagtggcccc acatctcaaa gaaaccctaa ataagaactt 10680

ttttgacttc ctccttacgt tcaaacagat tcatggggat acggtccaga accagctggt 10740

ggtccaagga gtggagctcc catcttacct ccccttgtac ccgcctgcca tggactggat 10800

cttccagtgc atctcctacc atgcccccga ggctctgctg accgagatga tggaaaggtg 10860

taagaaacta ggaaacaatg ccttgctgtt gaattctgtg atgtctgcct tccgggctga 10920

gttcatcgcc acaaggtcta tggatttcat tggcatgatt aaagagtgtg atgaatctgg 10980

tttccccaag catcttcttt ttcgatcact gggattaaac ttggccttgg ctgatcctcc 11040

tgagagtgac cgacttcaga ttctcaacga agcttggaaa gtcatcacta agctgaagaa 11100

cccacaggac tacattaatt gtgccgaagt gtgggtggaa tacacctgca agcatttcac 11160

gaaacgagag gtgaataccg ttttggcaga tgtcatcaag cacatgactc cagatcgtgc 11220

atttgaagat tcctaccccc agcttcagtt aataattaag aaagttattg cccacttcca 11280

tgacttctca gttcttttct cagtggaaaa atttctgccg tttctggaca tgttccaaaa 11340

agagagtgtg cgggtggagg tttgcaaatg catcatggac gcctttatca agcatcaaca 11400

agagcccacc aaggacccgg tcatcttgaa tgccctttg catgtttgca agaccatgca 11460

tgactctgtg aatgcactca ctcttgagga tgagaaaaga atgctgtcat atttgattaa 11520

tggatttata aaaatggttt cctttggccg tgattttgaa caacagctga gtttttatgt 11580

tgagtccagg tcgatgtttt gcaatctgga gcctgttctt gtgcagttga ttcatgcctg 11640

tgttgcctac tgcttcatca ccatcccctc cctggcgggc atcttcacac gtctcaatct 11700

ctacctgcat tctggtcagg tggccttggc caaccagtgc ctctcccaag ctgatgcttt 11760

tttcaaagcc gctataagcc ttgttccgga agttccaaag atgattaata ttgatgggaa 11820

gatgcggcca tcggaatcgt tccttctgga attcctctgc aatttctttt ctactttatt 11880

aatagttccg gatcatcctg aacatggggt cctgtttctt gttcgagagc ttctcaacgt 11940

gatccaggac tacacctggg aggacaacag cgatgagaaa atccgcatct acacctgcgt 12000

cctgcatctc ctctccgcca tgagccagga gacgtacctt taccacatag acaaagtgga 12060

ctccaacgac agcctctacg ggggagactc caagttcctg gcagaaaaca acaagctgtg 12120

tgagacggtg atggctcaga tcctagagca tctgaaaacc ctggccaagg acgaggccct 12180

gaagcgccag agctcgttgg gcctttcctt ctttaacagc atcttggccc atggggacct 12240

acgcaacaac aagctcaacc agctctccgt caacctgtgg cacctggcac agaggcacgg 12300

ctgtgcagac accaggacca tggtgaaaac gctagaatac atcaagaagc aaagcaaaca 12360

accagacatg actcatctga cggagctggc cctcagactc cctctgcaaa caaggacctg 12420

a 12421

<210> 129
<211> 1494
<212> DNA
<213> Homo sapien

<400> 129
cggccgccgg gcaggtacta ggttctgaca tcagtgaagt agcttccatt tcatttctct 60

gagtgcttca ctcagcaaga gactatgcga caactacgtg ctgggtgttc cagatgtatt 120

gaactctgga caaactgctt gacttcagcc tcttgaacta ttgttaatac acgggcagct 180

tgttatatag ctcccattgg tatacaactt gctaattaaa tggaaaatac ccttgggccc 240

acgtgccatt gtgcttgtgc tatacaacca tggtaaattc cccaggcatt ttattgggcg 300

cagtgtccac atgggccgtg tgttggtatt caacccttgg aagtgctcag gggaagattc 360

tgaagaaacc cattccttag tcgccaaagc gattggtttg acatgcccgc cggctctatt 420

ctaacctaac agacagaatt acggaagaca ttttatgtgc ttgttgtgtc cgctatgtgt 480

ttgtggctgc ggtgacaaac ccagttgcta aacacatacc ttccaaggcc tttacctttt 540

gcaacgcaat gtggcctttg cacacaagta ctggagaggg aagattaccg aactatatga 600

aaacccctct gtgcggaacg gtggcccatg gaacttggct atttggccac atggctagtt 660

aaggccacct aaaacggcaa aattttgggt tggtcgccta agcgtttggt caaatcagtt 720

gcccctaatg tggccctttg ggtgcacggt caactttatc cgccgaagtt tccgattcgt 780

aatagaatcg tagcatcatg acatgaaacg acacacacac acacaaaaca cactacaata 840

gcagcgctca ctgatgcagc gagcggtaaa taaccaccct ccgacagaga ggggcggacc 900

gaccacaaaa tagagacgca gtgtagagta ctacacacca cccaggttgt tagttgggta 960

gtgcagacac aacattcttc ttgtgaatat aaagcattcc cgcgagtgat ccttcctact 1020

accagtatta tcatccctac tcatgctatg gctacatcgt gcaacaatta ttcatgcacg 1080

tgcgagtgat agcgacacaa acatacggtg atacatctaa ctagacgcta acgcacatca 1140

gcatccaacg agaatatatt atgaacagat ggagcacgac tgaagtcgta ggttgattcg 1200

cacatgctgg acgtatgagt atccggagat gtcaactgac aggagtgcat gtgggcacag 1260

ggcttacggc agtagtagca ccatacggct gacctagaga tggattggta aatcaaagta 1320

gtgagggagc ggcggaatgg gaagcgtgac ggggatgtga tgttagctgt cggagttgga 1380

cggagggagg catgggagga acgtgtgagc aatagagccg gcgaggtgtc tacaggaagg 1440

gagagcagag gcgagcacgg agcgatcgct ggagctgcag ctgcccaggc ggcg 1494

<210> 130
<211> 1774
<212> DNA
<213> Homo sapien

<400> 130
tgtgagttcc acagaagatt tcagttgaaa agcactatta atgtttccta ctgtggacag 60

catttagagg aaacatttta agcagcacgt tcatatttcc tctgttcacg tacttagtca 120

acagctcttc attgtgcgct tatcgtgtgc aaggcagtgt tctacatgct atggggtaaa 180

caagatgaca aatccctgtc cccttagaaa agagcttgca accacagtgg ctgttggaaa 240

gaagcaacag tggtgctaaa agacaccagg taaactggaa aggaaaattc atcatatagc 300

gtactaggtt ctgacatcag tgaagtagct tccatttcat ttcttgaggc ttcatcagaa 360

gagatatgcg acaatacgtg ctgggtgttc cagatgtatt gaactctgga caaactgctt 420

gacttcagcc tcttgaacta ttgttaatac acgggcagct tgttatatag ctcccattgg 480

tatacaactt gctaattaaa tggaaaatac ccttgggccc acgtgccatt gtgcttgtgc 540

tatacaacca tggtaaattc cccaggcatt ttattgggcg cagtgtccac atgggccgtg 600

tgttggtatt caacccttgg aagtgctcag gggaagattc tgaagaaacc cattccttag 660

tcgccaaagc gattggtttg acatgcccgc cggctctatt ctaacctaac agacagaatt 720

acggaagaca ttttatgtgc ttgttgtgtc cgctatgtgt ttgtggctgc ggtgacaaac 780

ccagttgcta aacacatacc ttccaaggcc tttacctttt gcaacgcaat gtggcctttg 840

cacacaagta ctggagaggg aagattaccg aactatatga aaacccctct gtgcggaacg 900

gtggcccatg gaacttggct atttggccac atggctagtt aaggccacct aaaacggcaa 960

aattttgggt tggtcgccta agcgtttggt caaatcagtt gcccctaatg tggccctttg 1020

ggtgcacggt caactttatc cgccgaagtt tccgattcgt aatagaatcg tagcatcatg 1080

acatgaaacg acacacacac acacaaaaca cactacaata gcagcgctca ctgatgcagc 1140

gagcggtaaa taaccaccct ccgacagaga ggggcggacc gaccacaaaa tagagacgca 1200

gtgtagagta ctacacacca cccaggttgt tagttgggta gtgcagacac aacattcttc 1260

ttgtgaatat aaagcattcc cgcgagtgat ccttcctact accagtatta tcatccctac 1320

tcatgctatg gctacatcgt gcaacaatta ttcatgcacg tgcgagtgat agcgacacaa 1380

acatacggtg atacatctaa ctagacgcta acgcacatca gcatccaacg agaatatatt 1440

atgaacagat ggagcacgac tgaagtcgta ggttgattcg cacatgctgg acgtatgagt 1500

atccggagat gtcaactgac aggagtgcat gtgggcacag ggcttacggc agtagtagca 1560

ccatacggct gacctagaga tggattggta aatcaaagta gtgagggagc ggcggaatgg 1620

gaagcgtgac ggggatgtga tgttagctgt cggagttgga cggagggagg catgggagga 1680

acgtgtgagc aatagagccg gcgaggtgtc tacaggaagg gagagcagag gcgagcacgg 1740

agcgatcgct ggagctgcag ctgcccaggc ggcg 1774

<210> 131
<211> 531
<212> DNA
<213> Homo sapien

<400> 131
ccgaggtact taattctttg gcaacaaaca gcaagttttt atggtttaat tgtatttcct 60

ctctagagat cacaatactc tgggtatttt atatcccttc taaaggtcat ttgggttttc 120

aaatgggaag aatagtcaag ctaagctgga ctaaacctaa gtaatatttt atctcatcaa 180

aagaagttat taatctaact gggttagcat gagtcattca ttttttagac atgataaatg 240

ggaaacatat caaatcatta gataaattct gacctggaat tgaattcccc ctttcttaaa 300

atctttttaa tttgcttttt tcatagactc agtatggaac tcttaactga taagggagag 360

attcttgatc tggaaccgtt ccctgccatt ctccttttt ctctctgtct aggtagctgg 420

ttccatagtg cccgtcacga gggtccattt cagtttgatg acattagact gcttacactc 480

agctggatgc cttgctgttt gcagcaacat gattttacag tatgcttttc t 531

<210> 132
<211> 4309
<212> DNA
<213> Homo sapien

<400> 132
tggggcatga agttctgcca ctaaggcata accagtaccc aagtaggaaa ggattattga 60

tccctggagt aaagattcct agcttgcggg gatctcacta ttgatccccc ggagtaaaga 120

ttcctagctc gcaggaatct tacttgcaca cttaaggaaa ataccagagg gagccactag 180

caaccgaaaa atgaaggaaa gatttaattt ctctacccag gtacaaatcc aatgcatagt 240

atagtatatg ttatatgcag aaagggcact ggaggtgtga gagcccttg ggtaacagct 300

ctgctagttg ttaccttcac acttctcttt cttcttaatc tctggtttct cgtctgaaaa 360

atgagataat tggactagat taatagttct cagccttttc tagaacaaca aattagatta 420

tctctgaagt ccttttccaa aatttcatgc ttccagtctc tccccgtgat tgcatattat 480

ccctaggcta attaaaagtg tgctggttaa tcagtttctt tacttttttt tctctctttt 540

ttttattctg tgttcaggta tcatttgacc aacaacttgc tcttaatctt actgcgatga 600

ctgaaatctc caaagaggag agtatagtag ttataggtct agtaataata atcttgtaaa 660

caattcatta ttttttaaaa cttaatgctg tgtgcctaat cagtgttgag tatgactaaa 720

atgtaaatga gaaccaaata tgacactaac gctggcacta aattttattt ttacttgtgt 780

acttaaagta ctttactggc ctgttgggaa attgatttgt ataagattca cacctggttt 840

tggcagaaac agctttttaa agatacaatt aggctgactt aaaaatgttt tttattccac 900

agaatgtatc tttattatgt cattcttacc gtctctctgc ctttttttgac catcatgtaa 960

ctacgcagta gtaaatcatg gggtgcaggc tgatgaatta gtagtctttg taaaactaaa 1020

atgtaatgag cataggaaga gacagctctt ctggtgaaca tggcagtttt ccggtaatac 1080

cagccttgat gaataggcta aagacagctg tgtactatgg acattcaact tcactggatt 1140

gcctggctta gcttttcagt tattgtcttg ccgtgaaaca tggtgttgca ctgtacacat 1200

tttagaatag tggcaagaat gttctttaga actgcagata acccaacaga gaataccata 1260

ccaatatatc ggtttctcct gttacataga gatttggtat ttcagtgatg gttatggttt 1320

ttgactcacc tgtgtgtcag tcctgctaga gagacagtat agaatcaaga atttcttgat 1380

gtctttttaa aaatagagat tataatggcc ttgctagttg tcctgtgaag tgacagaccc 1440

ttaattagaa atcttttaat cctgctttcc attttacttc tccaccattt atttttaaca 1500

ttcattctct gaaaatagtt atgattttta gtgtaattgg ctatgtttaa gaaattccat 1560

ttcaaagtct taatatgtaa caatcttcat tttttataaa attgaaggac atttttgagg 1620

caaacttact tttatagctc attttctccc tagttaagga gaattctctg gtttcatata 1680

gagaatggct agctctgtga taccccaccc tctttgtggc ctaccagttg ctgttgtgtt 1740

gcttaaaatt gcagatattc tcataaacat gatgattttt ggcacaggcc tttctttgct 1800

attgatttca aaataaagtt gggcaatccg atttgaacta ctttaataaa cataatttag 1860

cattcctgta atgagaaagt attttcaaaa gataaccaga tttatttcta ggattagaga 1920

tggcaaatac caatgatagt atttttccca agagcaatca cagattaatc tataaatatg 1980

agaagtcgca ttattgtact taattctttg gcaacaaaca gcaagttttt attgtttaat 2040

tgtatttcct ctctagagat cacaatactc tgggtatttt atatcccttc taaaggtcat 2100

ttgggttttc aaatgggaag aatagtcaag ctaagctgga ctaaacctaa gtaatatttt 2160

ttctcatcaa aagaagttat taatctaact gggttagcat gagtcattca ttttttagac 2220

atgataaatg ggaaacatat caaatcatta gataaattct gacctggaat tgaattcccc 2280

ctttcttaaa atctttttaa tttgcttttt tcatagactc agtatggaac tcttaactga 2340

taaggagaga ttcttgatct ggaaccgttc cctgccattc tcctgttttt ctctctgtct 2400

aggtagctgg ttccatagtg cccgtcacga gggtccattt cagtttgatg acattagact 2460

gcttacactc agctggatgc cttgctgttt gcagcaacat gattttacag tatgcttttc 2520

tcaaagcttt gcattcttaa tggagatatc aaatggtgta attccaaata taaatatgtt 2580

tatgacacta atcatatgct tttaacaata acttttttgat aacttattgc cctgtaagtt 2640

aaaccttaca cggagtgact gttgccatca gaaaatcccc tttcacctgc tagagagagt 2700

gaaagtagtg cgtggcctcc acttttcagt taacgagtag caagcttcta gtaggggctg 2760

cttatctgac ccatgtgggg tcttgggcct tgttactttc ctgggttcgt cccttcagct 2820

ggaaaggctg ttgaaaacac ttgccaggaa acagtaaagc tgtgagagaa tcttctagtc 2880

ttagatatag agaaatgagt accagttgat gctaacacga tacttagact ttgaggggct 2940

tccatccaag ctttggcatg aatctgttag acacggtttt cttcactcct ctgataatct 3000

ttttttccag aatggttttt tgttgttgtt gttgttaaca aattctaaac atccagtgtt 3060

acttttttgtt ttgtttgttt tgttttgttt tgttttgttt tgtttgccct tcaggaaagt 3120

ccttttttcga gtaggtattg attgttccct ggaaactctc agcagtgtct gtgcaggctc 3180

tgtgcatgct ttgtatgagt tcggcttgaa caatgcattt gaagtcacct gggatgtcca 3240

gttctggcat gtcttcattg attgtgtttt taaacatgtt tcatgtttca tgtcattttc 3300

caaacctcac tttacatctt actcagagaa gttaatcaag gaatgaaatt tctaggggga 3360

tagggatgat gagggtgggt tgggggcttg agtgaagtca acttggggta tttgctttaa 3420

agtgttttct aaagcagttc ctaacagtat ttagagattc cctgggatgt ttgtggctca 3480

acttattctg gacaaagtgt gtgtgggggt ggtcttactg agatttgcat tcttaatgca 3540

ggacaggctc taaatttcat ctgtactcta aatttgtgat tgaatccaag aagataacag 3600

agacagtgct cctgttgtaa tgtttctggc aagtgctccc taaaatgcac atcgaattct 3660

gttttctggg ccttttctcc aatggtgcta ggagataccg ttgatttctg cagctcttct 3720

cagtggtggg aagaagtctt tgggattgtt gagcaagggg cagctggacc atccactaaa 3780

tttttttgtt caagacacat tagagaccct cctgtatatc tagtaagtca taataaaggt 3840

gcttgggaaa gccttaaatt tgaagacaca tggaggcggt agaaaattaa acttgtaaga 3900

ggagaaaaac atgccattag gtaacgcaga gttgtaacta ctggctaaga ctcaatggaa 3960

cttccacttg ctctaaaacc agggaaggga gtgggacgat aagtcttttg aagacatcag 4020

ctcactgtgc tgagagaggg accaaactca aggaaacctc tgatctatac attcaactgc 4080

tgcatttttt tataaataca tgtaaatgtc ctgttgtaat atttgatctt ggaaataaaa 4140

acaaaaactt ttcgtagaaa aaacacacac gacaaaaaca aaatttgggg ggccggcgcc 4200

aagaagtttc aaccacatgt gtggggccgg cccaggtgaa ggaaatggca agtggccaag 4260

aaccggccag aaggggggtt tcccaaaaaa cgcgcgggcg cgcgtttga 4309

<210> 133
<211> 730
<212> DNA
<213> Homo sapien

<400> 133
gccgccgggc aggtacaacg tggagatttc cctctgggaa aatcacccag caagtgaaat 60

ccgacccaga gggggacaca ggagtgctcc cgtgggacag cgtggccacc actgggaggg 120

gctctggcag tttttcttgt tttttacgct gcgatgtctg tgaatcttga gagcccagct 180

gatgcgttgt aacctcacct gggggcgatg tggaacattc caggcctggc aggggctatg 240

cctgccatgc agacctcacc cgagccatct caccctgggt ctgtcagagt cccacgtgct 300

gtcgcccctc acccaccccc gacggggcca tgctcctggt catgctgtga ttcattcatc 360

attccatggg caggtgtggg attgtctctg tgtttctgtc tgctttttaa agaagatgag 420

gtaagcatgg aaaacaaaac aaacgtggta accccatcac tcagacgtgt ccactgttag 480

cattcgggaa caggtgagag cagtctggcg cacaggtttt aaggctggct ggcccagact 540

ttggtcttgt gctgtgtgat cttggggaaa tcactgattt ctgagccttt ttcccagctg 600

ctttttctct tggggctggt agaaggaata agagagtgag tgtcaaaaaa aaaaaaaaaa 660

aaagcttggg gttaatcatt ggcaaatatg ttcccggtta aaattgtttc ggtcaaatcc 720

atttggagaa 730

<210> 134
<211> 226
<212> DNA
<213> Homo sapien

<400> 134
gcccgggcag gtactgggtt gcagcaagct atgatgccct gaggttccct gaagacaagt 60

aagaacacac tcaacttgtc actcagaaac gcagccctta aaggcttcca gataagaata 120

atgactgggt aaagccaggc tgccacctaa gctgaagcta caggtaggtg actaaggaaa 180

cactgatgtc aggccagcct aggccaactt ctccatgtgt tctttg 226

<210> 135
<211> 937
<212> DNA
<213> Homo sapien

<400> 135
ttttttttttg agatggagtc tcgctcaagt gcccaggcta gagacagtgg ctcaatctca 60

gctcactgca acctccgcct cctggattca aacgattctc ctgcctcggc cttctgagta 120

gctgggatta caggcatgcc ccaccacgcc cggctaattt ttgtattttt ggtagagatg 180

gggtttcacc atattgtcca ggctggtctc aaacccctga cctcaggtga tctggccacc 240

tcagcctttc aaagtgctga gattataggc gtgagccact gtgcccagcc gcagaaaagc 300

ttcttacctt gagcaggggc atctctcggg cctgctggat taaaaggtgc atttcgttga 360

tctgctgccg cacaaggggc ggtactgggt tgcagcaagc tatgatgccc tgaggttccc 420

tgaagacaag taagaacaca ctcaacttgt cactcagaaa cgcagccctt aaaggcttcc 480

agataagaat aatgactggg taaagccagg ctgccaccta agctgaagct acaggtaggt 540

gactaaggaa acactgatgt caggccagcc taggccaact tctccatgtg ttctttgcga 600

caccaagctt taagtggatc cctcttggag cctgactccc acaatcaggg agagctgaaa 660

gtaagtgcca cagcactgtg gattctcaat tccgcagggg gcaaccaacg ggctgtggat 720

tttcattctt gcttatactt caatcaagat cacaaaacgt ttaagatcat gagctgcttt 780

aatttggaag aaagtacaga tgacagaatc tgggacattt tcttcccttg caagtagatg 840

ctataaggct cagagactta aaataaactc aaactaaggc ataatcattg aaatgattcc 900

agttcagaag ttttagtgct taaatggcag cctgagg 937

<210> 136
<211> 96
<212> DNA
<213> Homo sapien

<400> 136
gcccgggcag gtaccaatca attttgaaaa tgaaagggaa aatacttatc tttcctatat 60

gaattgggaa aataaaagga tatgaaaagg taaata 96

<210> 137
<211> 151
<212> DNA

<213> Homo sapien

<400> 137
aaaatgtagg cagcaaaagt ggaagaggag aggcagctgg tgcactaatc caggtgagag 60

gtaaagatgg atgggtaggg ctgaaggatg gtggggcagg tggtgagaag tgacttggtt 120

ctggagacat atgaaggaag atggtcaggc a 151

<210> 138
<211> 604
<212> DNA
<213> Homo sapien

<400> 138
cgtggtcgcg gcgaggtact ttcttagacc agtgtaacct cacacctcag tttggctttt 60

ccaaccctga cttgaaaggc atatttgtat ctttttatta gtgatagtga agctgtgaca 120

ctaacctttt atacaaaaga gtaaagaaag aaaaactaca gcgattaaga tgagaacagt 180

tctgcagttg ttgaactaga tcacagcatt gtaggcagaa taaaaaatgt tcatatctga 240

gaatattcct ttcgccatct tttcccaagg ccagacctcc tggtggagca cagttaaaag 300

taacattctg ggccttgtaa tcggagggct gtgtctccag ctggcagcct tgttttaata 360

tataatgcag gactgtggaa aacagttggc atagaatatt ttcacctaaa aaagaaagaa 420

aagacataca aaactggatt aattgcaaaa agagaataca gtaaaatacc atataactgg 480

acaaagctag aagaaccttt agaagatttg tctgaaaaca gatttcaaga gtgagctttt 540

atacactgct cactaatttg cttgattact accaactctt cttaaagtta acacgtttaa 600

ataa 604

<210> 139
<211> 4461
<212> DNA
<213> Homo sapien

<400> 139
tgagttttcc cctgttggcc agggatggtc tcgatctcct gacctcgtga tctgtccacc 60

tcggcttccc caagtgctgg gattacaggc atgagccacc gcgcccagcc tagccatatt 120

tttatctgca tatatcagaa tgtttctctc ctttgaactt attaaccaaa aaggaacatg 180

cttttcatac ctagagtcct aatttcttca tcatgaaggt tgctattcaa attgatccat 240

cattttaatt ttaccaatgg ctcaaaaatt ctgttcagta aatgtctttg tgactggcca 300

atggcataaa ttatgtttaa gattatgaac ttttctgaca gttgcagccc atgttttccc 360

tacgatacca gatttccatc ttggggcata ttggattgtt gtatttaaga ccgtcagaat 420

aatgatagtg tgtggtctcc agaggtagtc agaatcctgc tattgagttc tttttatatc 480

ttccttttcc attttttatt accattttgt ttgtttagac tacactttgt agggattgag 540

gggcaaatta tctcttggag tggaattcct gtgttttgag ccttacaacc aggaaatatg 600

agctatacta gatagcctca tgatagcatt tacgataaga acttatctcg tgtgttcatg 660

taatttttg agtaggaact gttttatctt gaatattgta gctaactata tatagcagaa 720

ctgcctcagt ctttttaaga aggaaataaa taatatatgt gtatgaattt atatatacat 780

atacactcat agacaaactt aacagttggg gtcattctaa cagttaaaac aattgttcca 840

ttgtttaaat ctcagatcct ggtaaaatgt tcttaatttg tctgtgtaca ttttcctttc 900

atggacagac cattggagta cattaatttt cttaatctgc catttggcag ttcatttaat 960

ataccatttt ttggcaactt ggtaactaag aatcacagcc aaaatttgtt aacatcaaag 1020

aaagctctgc catatacccc gttactaaat tattatacat ccagcagatt ctgggatgta 1080

ctaacttagg gttaactttg ttgttgttga taatactaga ttgctccctc tttaattctt 1140

cttctggtgc aaggttgctg cttaagttac cctgggaaat actactacaa ggtcaaattt 1200

tctagtatct tacagcctga ttgaaggtga ttcagatctt tgctcaatat aaatggattt 1260

tccaagattc tctgggccat ccttgaccca caggtgatct cgctggagta tattaactta 1320

acttcagtgc cagttggttt ggtgccatga gatccataat gaatccagaa cttcaccatt 1380

gcttagatat aagagtccct tggaagaata atgccactga tgatgggggt cagaaggtgt 1440

attaactcaa catagagggc ttttagattt ttcttcaaaa aaatttcgag aaaagtattc 1500

ttttaccctc caaacagtta acagctctta gtttctccaa atatgctctt tgatttactt 1560

atttttagtt aaagatggta atttattgaa caatgaaatc cgtaatatat tgatttaagg 1620

acaaaagtga agttttagaa ttataaaagt acttaaatat tatatatttt ccatttcata 1680

attgttttcc tttctctgtg gctttaaagt ttttgactat tttacaatgt taatcactag 1740

gtaacttgcc atatttctgg ttctatatta agttctatcc tttataatgc tgttattata 1800

aagctggttt ttagcatttg tctgtagcaa tagaaatttt actaagtctc tgttctccca 1860

gtaagttttt tcttttctca gtaagtccct aagaaaacat ttgtttgcca ctcttactat 1920

tcccaatctt ggattgttcg agctgaaaaa aaatttgatg agaaacagga ggatcctttt 1980

ctggtgaata taggttcctg ctttaagaat gtggaaatcc attgctttat ataactaata 2040

tacacacaga ttaattaaaa ttgtgagaaa taattcacac atgacaagta ggtaacatgc 2100

atgagttttg aatttttta aaaacccaac tgtttgacaa aatatagaac ccaaattggt 2160

actttcttag accagtgtaa cctcacacct cagttttgct tttccaaccc tgacttgaaa 2220

ggcatatttg tatcttttta ttagtgatag tgaagctgtg acactaacct tttatacaaa 2280

agagtaaaga aagaaaaact acagcgatta agatgagaac agttctgcag ttgttgaact 2340

agatcacagc attgtaggca gaataaaaaa tgttcatatc tgagaatatt cctttcgcca 2400

tcttttccca aggccagacc tcctggtgga gcacagttaa aagtaacatt ctgggccttt 2460

gtaatcggag ggctgtgtct ccagctggca gcctttgttt taatatataa tgcaggactg 2520

tggaaaacag ttggcataga atattttcac ctaaaaaaga aaaaaaagac atacaaaact 2580

ggattaattg caaaaagaga atacagtaaa ataccatata actggacaaa gctagaagaa 2640

cctttagaag atttgtctga aaacagattt caagagtgag cttttataca ctgctcacta 2700

atttgcttga ttactaccaa ctcttcttaa agttaacacg tttaaataag gtatttctgg 2760

acttcctagc cttttagcaa gcttagagga actagccatt agctagtgat gtaaaaatat 2820

tttggggact gatgccctta aaggttatgc ccttgaaagt tcttaccttt tctctagtga 2880

tattaaggaa cgagtgggta gtgttctcag ggtgaccagc tgccctaaag tgcctgggat 2940

tgagggtttc cctggatgcg ggactttccc tggatacaaa acttttagca gagttttgta 3000

tatatgtgga tttttctgat aagtagcaca tcagaggcct taaccactgc ccaaaagcga 3060

ttctccattg agagtacata tcttgaactt aagaaattca tttgctctga tttttaatct 3120

tgtaaagttt ttgctaaact caaaacaagt cccaggcaca ccagaaggag ctgaccacct 3180

taggtgttct tgtgatttat ccttacttcc ctatgttgtc atagttgctt ctaaactcag 3240

ctgcactatg gctgtcaaca tttctgatac ttattgggat atgtgccatc cagtcattta 3300

gtactttgaa tggaacatga gatttataac acaggtaata gctgaaggta ccagtatggt 3360

ggtgagactc acacttagtg atccagctaa ggtaactgat gttataatgg aacagagaag 3420

aggccaacta gatagctaag ttcttctgaa cctatgtgta tatgtaagta caaatcatgc 3480

gtccttatgg ggttaaactt aatctgaaat ttacattttt catagtaaaa ggaaaccaat 3540

tgttgcagat ttcttttctt gtgaggaaat acatggcctt tgatgctctg gcgtctactg 3600

catttcccag tctgttctgc tcgagaagcc agaatgtgtt gttaacattt ttccgtgaat 3660

gttgtgttaa aatgattaaa tgcatcagcc aatggcaagt gaaggaattg ggtgtcctga 3720

tgcagactga gcagtttctc tcaattgtag cctcatactc ataaggtgct taccagctag 3780

aacattgagc acgtgaggtg agatttttt tctctgatgg cattaacttt gtaatgcaat 3840

atgatggatg cagaccctgt tcttgtttcc ctctggaagt ccttagtggc tgcatccttg 3900

gtgcactgtg atggagatat taaatgtgtt ctttgtgagc tttcgttcta tgattgtcaa 3960

aagtacgatg tggttccttt tttattttta ttaaacaatg agctgaggct ttattacagc 4020

tggttttcaa gttaaaattg ttgaatactg atgtctttct cccacctaca ccaaatattt 4080

tagtctattt aaagtacaaa aaaagttctg cttaagaaaa cattgcttac atgtcctgtg 4140

atttctggtc aatttttata tatatttgtg tgcatcatct gtatgtgctt tcacttttta 4200

ccttgtttgc tcttacctgt gttaacagcc ctgtcaccgt tgaaaggtgg acagttttcc 4260

tagcattaaa agaaagccat ttgagttgtt taccatgtta ctatgggact aatttttaat 4320

tgttttaatt tttatttaaa ctgatctttt tttatatggg attacatttt ggtgttcact 4380

ccctaaatta tatggaaacc aaaaaaagtg attgtatttc acatatggaa aaaaaaaaaa 4440

aaaaaaaaaa aatatgcggc c 4461

<210> 140
<211> 321
<212> DNA
<213> Homo sapien

<400> 140
cgaccggcgc tttgtgatgg atccgcccgg gcaggtactc tggtagccca gtttgcaaga 60

gccaggttcc agcagggatg gtagccagcc tgggcacact tttctgtatt gtatttgaac 120

ctcatttcct gatgctcatc cagaagttcc tggctgggtg agtaggtgaa cagttcaagg 180

ctgtgtggat gaaactctgg catctcttta cagtgctgtc cggcagggca gtgtgaggcc 240

actctgtcct cctttgacct ctttcttctt cagcagccaa atggaggctc tctaaaggtg 300

atgattaact tggcataaac a 321

<210> 141
<211> 1438
<212> DNA
<213> Homo sapien

<400> 141
gggggacaca gcactgcctg aagcagcatc ttgtaggtca gccagttccc gttgacacgc 60

agaaagacag tactgctggg gttcattttc tgggaaagcc aggcagagcc tgaggaagct 120

tcatgggaag gcagcatgga acttgagctg agccacggag gatggggaag tgggagaagg 180

tcatcgaggt agctggaaag gagagtgttc ccagggagca ttttagaggt gggaacattc 240

aggtgcatag aggtgagagt ggagatgctc cagggagttt gacagcaagc acttggcagt 300

gatacagggg caggggtaca gtctgaaagt gctggactgg ggtactctgg tagccaagtt 360

ttgcaagagc cagggtttcc agcagggatg gtagccagcc tgggcacact tttctgtatt 420

gtatttgaac cccatttcct gatgctcatc cagaagttcc tggctgggtg agtaggtgaa 480

cagttcaagg ctgtgtggat gaaactctgg catctcttta cagtgctgtc cggcagggca 540

gtgtgaggcc actctgtcct cctttgacct ctttcttctt cagcagccaa atggaggctc 600

tctaaaggtg atgattaact tggcataaac atgccaagct ttcctgagcg tttgccctga 660

attgggctgt ctcttgctgc tctgagacag ggagactggc tgtctggctg accagccctt 720

ggggctgagg tctggctgtc aagttgcagg ttccaggcaa ccagcaattc ctggttgtca 780

cctggctcta gagattagag cagcagtgtt gctcactaga actttgtgca gtgatagtgt 840

tctttaatct gcactgtcta gtactgtagc catgggccac ctgtgacttt caagcccttg 900

aaatgtggct accggccagg cacagtggct catgcctgtt atccaggggt gtccaatctt 960

ttggcttccc tgggccacaa tggaagaaga attgtcttgg gccacacata atatacagta 1020

acactaatga tagctgatga gctttaaaaa aaaaatcaca aaaaatatct cataatgttt 1080

taggaaagtt tacgaatttg tgttgggctg catttgaagt ggtcctgggc cttggtctgc 1140

gggtgggaca aagcttgtta taatcccagc attttgggag gtggaggcag gacgatcact 1200

tgggcccagg agtttgagac caggctgggc aatatggcgg atcccatctc cacaaaaatt 1260

gcaaagttac cgggatggtg tcaggcgcta tagtccagct gcttgggaga ctgagcagga 1320

agaatgctta gctgtaggga cagttgcggg cggatacagt atggcccagc caggcaaggt 1380

gacccgtttc aaaaaaacaa cattggccgg cttataggta ataaacacaa acgatgta 1438

<210> 142
<211> 368
<212> DNA
<213> Homo sapien

<400> 142
ggccgaggtt agcatttaat ctactggaat aagctggttg ctgtggctta tacctgtagt 60

cccagatact tgggaggctg aggcaggaga atcacttgag accagcctgg gcaatatagt 120

gagacctcac ctctaaaaaa gtttttttata aatttacttg tctaaagtgg ggaaagggaa 180

attattctgt tttcttattc ttgcttcaag actatgacag atttgaaaga gaattctaaa 240

gcagatttag agaacctgct tctcttctta tctcctaatc cctaaattgt aatttagctc 300

ttatctgtat tgtgttttgt tttggtaaag ggatgatttt tacattgagt tttaaagtag 360

aataagaa 368

<210> 143
<211> 540
<212> DNA
<213> Homo sapien

<400> 143
gccgcccggg caggtactaa gcatattaca ccagtgtgtg tggcttcaaa tagaagactc 60

tctatctcta ggtatcgaaa tggcgaataa gaattgcagc tctgaaaata taatataaac 120

acatttttca gatagcatta tacaaataaa aatacttttg gatactagaa tttggggtta 180

gagctgactt caataagggg aggaaaaaat atgtgcaaag ggaagggagc tcaatttgga 240

ataatatatt tcaaagtatg tggctttgca tttttaaaat gctaaccaaa aaaaatgtga 300

gatagagtga ccccaaggaa ctttttacag ccagcctctg gataccagat atagcattta 360

caacttctga tatgctagac aaattttcca tagttttaat tttcttaaag acatatagtg 420

agtgaacacc aacactggga aggaggtagt gttaatttga tcaaatattg atatttcccc 480

actgattcag aaagtgcttg ctctttctct aggatatctc aaaacccttc caatttttct 540

<210> 144
<211> 1195
<212> DNA
<213> Homo sapien

<400> 144
aaagactgca tctatggccc accttctcat aacctcaaac gcagagtact cattctaaga 60

aacacatgca ttcttgcctt aactgaggac tgaaatacca aacccaacca aaaataacag 120

aacttttaag tgccctgatg gaaatgtaca gtatgagtta catccatcca tctatacaca 180

catacagaca ctcatatgta tatacaatat atacatacaa atgataatag ctataccaaa 240

aatgtgcata tttaatacag ttttaaaaag ccagtagatc atggtataat acaatttatt 300

tccactagtg aatggcacaa atggagaaat gcaaaggagg tgtttcaagt acaatgaagt 360

catttttcac aggtgaaaat aaatagtatt attccagtgt ccaacataat gtattctaca 420

aaagtttaaa atgccagaca tttagtactt acagaaatag aaagttacta aagtctgctt 480

tatggtctca aaaagacact catagagttt tcaatttttg aaaaggggggt gagaaaaaaa 540

ttggtggaag caggagggag tgtccaggat agtttcttta gtgtactgaa tgtttacctg 600

atgtctaggc aaccaaaaat catgactatg ctaaaatgtt ccatttacca acaaatcctt 660

gacctagagt actaagcata ttacaccagt gtgtgtggct tcaaatacaa gactctctat 720

ctctaggtat tgaaatggcg aataagaatt gcagctctga aaatataata taaacacatt 780

tttcagatag cattatacaa ataaaaatac ttttggatac tagaatttgg ggttagagct 840

gacttcaata aggggaggaa aaaatatgtg caaagggaag ggagctcaat ttggaataat 900

atatttcaaa gtatgtggct ttgcattttt aaaatgctaa ccaaaaaaaa tgtgagatag 960

agtgacccca aggaactttt tacagccagc ctctggatac cagatatagc atttacaact 1020

tctgatatgc tagacaaatt ttccatagtt ttaattttct taaagacata tagtgagtga 1080

acaccaacac tgggaaggag gtagtgttaa tttgatcaaa tattgatatt tccccactga 1140

ttcagaaagt gcttgctctt tctctaggat atctcaaaac ccttccaatt tttct 1195

<210> 145
<211> 787
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (344)..(634)
<223> a, c, g or t

<400> 145
acatttagag aaggttaaaa gaaacagtga gaaatgtaaa cattcaaaat gataattgaa 60

tctctcagtt gtgggaataa ttatcagaga catgcaactg aaaatgtctc acctttcatc 120

ttttttttctt aattcataaa gttatcttgt agaatttgat gagaccctcc tagtcattct 180

caactggggc ggtgctgtca ccgaatggtg tttgagagtg ttggggctag ggcacatttt 240

tggttgtcac agcaactggg gtggcatttg ctgcccagtg ccaggaatag taacattatg 300

aatgccagga cagtgtgctc agtaaagtct tccatccaaa aggnnnnnnn nnnnnnnnnn 360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnncgaaaa catcccgtaa aaaaaaacac 660

ccccggttcg cttgtccctt ccccccttgtg acaaactttt tcaaatccag aaagctgcgg 720

gtaatcatgg gcatagggct gttcctgtgg taaaattgtt ttccggccaa tccccaaaac 780

aaaaaaa 787

<210> 146
<211> 193
<212> DNA
<213> Homo sapien

<400> 146
tggtcgcggc gaggtacagg ccgcctactt actgccctgc ctgacccagg attcttagga 60

ggggaaagtg ctatccctgg ggccaggcag ggagccacag ccactcactg cattctcaga 120

attctatcct accaccccaa agattgacct gaaagagact ccccttcctc agtcatgtag 180

aagtccttta cat 193

<210> 147
<211> 661
<212> DNA
<213> Homo sapien

<400> 147
tcatgccgag cgcgcgcatg agtgatggat tggtcgcggc cgaggtacag agtcctgtta 60

tttttctctt tggccctatt tggctgctta tattaatgca tcagaacttt atgtataatc 120

atatggattt atacgtaaat taagaaaaaa tgtccatttc attcagttca tatgttctaa 180

acgtattgct gatcattctt aaatgagact ccaggtttac attcttacat aaagtgcagg 240

gatcccgaag ttagccccaa agatcccctt gccttttttca gacttgctca aatgttacct 300

tatcagtggg gcctttcctg accacacttt aaaaacctca acacccaccc atgggccttg 360

tcctccttcc cggcttcatt ttttggcata tacttatcaa atgtgaacat atgatgcatt 420

tgctttattt atcatcgatc ttcactcact ggcatgtaag ctctgtgagt gcaaagattt 480

tcatctagct aatcttccag aacagtgtct ggcacagaga aggagctcta tgaatatgtg 540

ttgaatgaat gactatcttt gccttgtaaa ccccatgcta ttggctctct cttcaggtgg 600

ctgaccactg caccccaggg catgctggaa agacaggagt cccaagccct cccttctgct 660

c 661

<210> 148
<211> 1897
<212> DNA
<213> Homo sapien

<400> 148
agttttgcgt tgcgccttgt tgcttgcgtc cgtcgtttgt ttgcctgtgg cttctgccgt 60

ctttttgtgg gctcgccgtt gcccgtgcct gccgctgtca cccttggcgc ccgcctgggt 120

gctggggtcc gcgattgcag tcctttgata gtgttagtga gggggggctgt cgtgcgtgtt 180

gtgtatggtc cgcatggggg agtcattagc atgttgagtt gactgtctcc cggtccgttt 240

aacgtgcgtc tggaaggtac atttttgtaa atcaagtagt tggaactaaa tccaacactg 300

ataattgcca tttcaaacac tgatctgaaa agtgaattag aagctgtaca atatcatcat 360

tagaaattct gcatatggct aataaatatt ccttttaaaa ttaatagagt ctaaagtctt 420

ccacatgatc tttacagata gagtgggaca ctatagaatt ctgattatat gatttagatt 480

ttagggatgt tttaacattt tcaaaccact agaaggacat tgggaacaga aagtaataga 540

gccaacgtca cgtggtaatg atcaatagtc cagttctacg aggagaacaa ttttaagctc 600

ttcactgagg ccaattctgc tgtattctaa ttccttttag gttcttggtg gtagagtaat 660

gagctatgac catctctgga atactggtga ggaaaatggc agcagtaaag aaatgaggaa 720

aatattacct aattaatgat aaagttaggt ccagtacaga gtcctgttat ttttctcttt 780

ggccctattt ggctgctttt attaatgcat cagaacttta tgtataatca tatggattta 840

tacgtaaatt aagaaaaaat gtccatttca ttcagttcat atgttctaaa cgtattgctg 900

atcattctta aatgagactc caggtttaca ttcttacata aagtgcaggg atcccgaagt 960

tagccccaaa gatccccttg ccttttttcag acttgctcaa atgttacctt atcagtgggg 1020

cctttcctga ccacacttta aaaacctcaa cacccaccca tgggccttgt cctccttccc 1080

ggcttcattt tttggcatat acttatcaaa tgtgaacata tgatgcattt gctttattta 1140

tcatcgatct tcactcactg gcatgtaagc tctgtgagtg caaagatttt catctagcta 1200

tcttccagaa cagtgtctgg cacagagaag gagctctatg aatatgtgtt gaatgaatga 1260

ctatctttgc cttgtaaacc ccatgctatt ggctctctct tcaggtggct gaccactgca 1320

ccccagggca tgctggaaag gacaggagtc ccaagccctc ccttctgctc tactccaagc 1380

ttttcttctt gggtgcattg actcaagtca ggtagtactt ctctatgtct gagcacagac 1440

gggctgtgtt catgtatttg tacatatgtg tgaatagaca gagaaactag tagcatgggt 1500

atgtgggggа atccatcttt tagggagaga tttatctact gtttttgtgt ttagtctcac 1560

ctcagaccag gttaagctgg ccagggctca tagttttcaa agagcaacag aaaaaatctg 1620

tttagcttac attctaagca tgttctcttt atctttcctg aaagctatcc acttttaatt 1680

tcatctcata ctacagagaa aatattattt gaaactgata gctttccaga aggttactga 1740

aatcacttat ttttcagtgt cttcactggc accattcata gtagctaaca ttagccactt 1800

tccgtgggcc tggtgctgtg ttaaagtgct ttacatatat tatttctttt aatccgcaca 1860

atgatccttt caagtaggta ctgttattat tcccaat 1897

<210> 149
<211> 254
<212> DNA
<213> Homo sapien

<400> 149
ccgaggtacc catctagctt ctggggctcc actgacagct gaggacagtc cacacccgac 60

ctggacccca ccccaccctg ggtctgtcca tctcagtccc ggcctgagcc tctgggccaa 120

agccacctct tctgagcagg caggcagagc gaaagactgg gagcagcaga caggggcaga 180

gcacggccca tgagcccacc ctccacttcc cagattggtc agagttacat ggtcaccttc 240

cctgcacctg cacc 254

<210> 150
<211> 1993
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1822)..(1822)
<223> a, c, g or t

<400> 150
aaggatcctt aattaaatta atccccccc cccggggcgg cccagcggat cgtgccgcgg 60

cggccgagcg cagctacagg agggtgtcca gaagccacaa gccatggctg tggggaacat 120

caacgagctg cccgagaaca tcctgctgga gctgttcacg cacgtgcccg cccgccagct 180

gctgctgaac tgccgcctgg tctgcagcct ctggcgggac ctcatcgacc tcgtgaccct 240

ctggaaacgc aagtgcctgc gagagggctt catcactgag gactgggacc agcccgtggc 300

cgactggaag atcttctact tcttacggag cctgcacagg aacctcctgc acaacccgtg 360

cgctgaagag gggttcgagt tctggagcct ggatgtgaat ggaggcgatg agtggaaggt 420

ggaggatctc tctcgagacc agaggaagga attccccaat gaccaggttc gcagccaggc 480

cagattgcgg gtccaagtac cagctgtgcg ttcagctcct gtcgtccgcg cacgcgcctc 540

tggggacctt ccagccagac ccggcgacca tccagcagaa gagcgatgcc aagtggaggg 600

aggtctccca cacattctcc aactacccgc ccggcgtccg ctacatctgg tttcagcacg 660

gcggcgtgga cactcattac tgggccggct ggtacggccc gagggtcacc aacagcagca 720

tcaccatcgg gcccccgctg ccctgacacc ccctgagccc ccatctgctg aaccctgact 780

ggtaaacaac tgctgtcaga aaagggctgg gcttgggaag gggaggtgga ggccaggtgt 840

ccccagacct ctaacccttg cccctagcag cctcttcttt gtggagcctc tcagtgtggg 900

cagccctcgc atgctggggt cgggccagct ctccccgaaa ggtcttgacc tgaatgatgg 960

ccggggaagc ctgcgtgtgc ccctttcaga gacggagcac ctgagatgtg ggaggtgcag 1020

catgttcccc tgggcccctc agaaagtcga gcttggaggc cagcctggat ctgtctctcc 1080

cttcccctcc tgggaccatt ctacctgtgt tctttgaccc tcggagcagg gacaggcaag 1140

acaactggca agcttgcagc tgccctgatg gtgcaggtgc agggaggtga ccatgtaact 1200

ctgaccaatc tgggaagtgg agggtgggct catgggccgt gctctgcccc tgtctgctgc 1260

tcccagtctt tcgctctgcc tgcctgctca gaagaggtgg ctttggccca gaggctcagg 1320

ccgggactga gatggacaga cccagggtgg ggtggggtcc aggtcgggtg tggactgtcc 1380

tcactgtcag tggagcccca gaagctagat gggtaccagg tggggttagg ttcccagagg 1440

actgagggaa tcctgtacag gatgtcccag ggtagatggg gagcaggatt gggacctgct 1500

ctgacagctg gacacatgag ccctggatga gtatggtagg gggtttgaag aatcccctgt 1560

ccacctccca aatccaggcc cggccccctc tggcttggag agcattccaa gcccccaccc 1620

caccctaga actgccattc ccaagacctc tgtctcccag ccaaccaccc ttggaacttg 1680

cctcttgtcc tgctggaaag atagcagtgt tctcctgact tcgccctact gcatgcagcc 1740

aaataaaagg tgtgcccagt ctaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1800

aaaaaaaaaa aaaaaaaaaa anaaaaaaaa aaaacaaaac tattgcttgt gtttccgttt 1860

ggattggctt tgcgtatcta ctttcagtgc tttcgtcatt tccttcttcc tagctgctta 1920

tcatgcgata ggctacactg tgctcctttc tcaattttac ccagttctaa tgggaatggt 1980

gtgtttagat ttt 1993

<210> 151
<211> 170
<212> DNA
<213> Homo sapien

<400> 151
gcgtggcgcg gccgaggtac atactcatat ttataaggac ttcctcacta ggagagattc 60

ctgggttcta cagaataaaa ttcttggcta cttgtcttat agctctgaac agacttattt 120

tcccagagag tatgtttatt atgtaatagc gagttgcctg acccccaaa 170

<210> 152
<211> 1394
<212> DNA
<213> Homo sapien

<400> 152
taggtctgta tagaatacga ctgaaagatt ttgactgttg atttcctttg aagagattct 60

cctgttctcc tcttcagagg aaagagagag gagagaggga gagaggagag aaaagagaga 120

agagagggtg acttttcatt atatgccttt gtaatattta atctcctccc ttttctcacc 180

cccttgagca tgtatgcttt tcaacaatta aaaagattaa aaagctctta aatagataag 240

aatttcttag agcaactagt aagtacacta actgccaatt gttttgaaat gtagacattc 300

tgaaaaatca aaatgattgt gccgaatgta ttttaaaagg ctaaaatatt atagcatttc 360

aaagtacata ctcatattta taaggacttc ctcactagga agattcctgg gttctacaga 420

ataaaattct tggctacttg tcttatagct ctgaacagac ttattttccc agagagtatg 480

tttattatgt aatagcgagt tgcctgaccc cccaaaaagc tgtgttctat catattaaaa 540

taaggcaaaa tgattacttt cagattaaga aattgtggga ctctagatct tgttatatag 600

tgaagttctt taaaaaactg aggtcttggt tctgaataat agtgggttta cattaattta 660

tttagaattg tcattggggg tatctctgac ctatttttat aaaataatct caatttttaa 720

aataggagta aaatgctcat tggcataagc cagtaataat aatttagtat ttttccaagt 780

atttatagtc aatgtgttgc catgaacttt tttaagggat gtttttaatt ttagaagtgc 840

tttaaaaagc aatatggcat ctggctctgt agaagtagaa aacatggtaa cttcaatgtg 900

atatatttgc ttttttcccc tcttaggtct ttggggtaaa aaaaatccaa agtttactca 960

attttatttc tacatatatt acctacaaat tatagaggtg agacctgctt gctgcctgtt 1020

catacctgtc agtgatacta atacctcccg tctgtacagt gtttcatagt ttccaaaggg 1080

ttctcgcatg gttcctcact tgagcctggt gaggaaatac ctgtaaacat gggtagaagt 1140

tttagtccct tgtaaaactc tggtataata tcaaaaccag gaaattgttc acagattcta 1200

aggattgggg aaaagagaaa aataaacaaa ttgctcagga gtagaattga aaaaaagaaa 1260

aaaaaaaaaa aaaatttggg ggcgggccga aatttaaaca gttgggggggg ccaggtggaa 1320

tggaagggca ggaaacggga gatggtgttc caaaaaagcg gggcgggtga aaaagggcgg 1380

gggggtacga gggt 1394

<210> 153
<211> 368
<212> DNA
<213> Homo sapien

<400> 153
gcactgagaa aggatatgga caagtcagtc agcattcaca attaagagaa aaacatctgt 60

gctttggaaa atgttcttca aggatagaga attgtgccct atgtccacca aatttgcatg 120

agatctttat aagattagac agccagtgga taaggcccct tatctttctt catggatggc 180

tgaggaaatt ctccgccttc cctgacatca gctgcataac tgtatttctg cctcgtggaa 240

ataaagtaga tgatcaggca cttgcggttt gttcttaata caagaaagac aatttgattt 300

ttaaaagttt tgatttgtag aataatgtaa gacaatatgt ttctttctac tttggttttt 360

ccattcaa 368

<210> 154
<211> 864
<212> DNA
<213> Homo sapien

<400> 154
ttgatatatt gcattcttga gcattaggct tctaggtgat ttgttaaact catagcaggt 60

tttagtacac agtgctgttt atgacagaaa aaaattttat cctacctctg aaataattgt 120

actttctgtg attcagataa aaactttata gaaactccct aatgaaaata ttgaagcatt 180

aaccagaaaa tgagtcagct ttttgtttcc aaaatgatgc aacaggaaaa cctttaacta 240

cttataatcc cgtatagtca ccatcaccac aaagtattga aaatctgttt tctcttttac 300

taagtgtctg cacggtcact tatgtatacc caaagccaga aagatatttt tatctcaggg 360

aaattccaga aatggaaaca tttttgtgta atattgattc atttctgtct caccaaagat 420

gtgttttcca cgtagcaaag aacatcagcc ccacgttata gggaacaagc gagtcccaaa 480

tcgtaccatc tgctgagcac tgagaaagga tatggacaag tcagtcagca ttcacaatta 540

agagaaaaac atctgtgctt tggaaaatgt tcttcaagga tagagaattg tgccctatgt 600

ccaccaaatt tgcatgagat ctttataaga ttagacagcc agtggataag gccccttatc 660

tttcttcatg gatggctgag gaaattctcc gccttccctg acatcagctg cataactgta 720

tttctgcctc gtggaaataa agtagatgat caggcacttg cggtttgttc ttaatacaag 780

aaagacaatt tgatttttaa aagttttgat ttgtagaata atgtaagaca atatgtttct 840

ttctactttg gtttttccat tcaa 864

<210> 155
<211> 179
<212> DNA
<213> Homo sapien

<400> 155
gcggcgcggt gttatggagt agcgtggttc gcggccgagg tacatgtttt taaaaaatga 60

ctacatgttt cacctggtcc tatttttgct atttggacca tacttttaag atgaattgat 120

cttacataca tgttaagtct gatttatctc cccacatttt taaacactaa atgaagctt 179

<210> 156
<211> 1849
<212> DNA
<213> Homo sapien

<400> 156
gcttgatacg ctcctaagga atttgccctc gagcaagcaa ttcggcacga ggctcgaacc 60

cctgacctca agtgatcagc ccgcctcagc ttcccagagt gctgggatta caggtgtgat 120

ccactgcacc cggccggcat tatgattttg tgtactcttg aaatggttat ctttgtggat 180

gatttttttt ttttaagctg aaacttacct catgaataac ttgattaaag tagtaggtga 240

ttaaaatttc aatagaatca aatgagacaa aaattttaaa ctgactcatt tgagtttcaa 300

ctttacagtc attgaccata aagcacacta aaaatgtaag ttatttttaa atacatctga 360

aataaaaata cttactaaaa aggaagaagc cgaagatgta tatttagacc agcacacaat 420

tttgatttca attagcctta ttctaatatt tagcttttag atctttcata cacattttca 480

cgtactttgc aattgagacc agaaagactt gtaggtcttt ctgcagaatg agtgggtcct 540

tgcaaagtga gtgggaaact tactcctaga tcagaaatgt ttgcctctct gagtaaaatg 600

tttctttcag atgagccata gaggggcac cttttactca acttttcttt gttttgaaac 660

tttgtttccc atactgtttt cagccttttg tttataatta gaaattgtga gaagcttcat 720

ttagtgttta aaaatgtggg gagataaatc agacttaaca tgtatgtaag atcaattcac 780

ttaaaagtat ggtccaaata gcaaaaatag gaccaggtga aacatgtagt catttttaa 840

aaacatgtac ttggtctttt gtgtgtgtct gttttattcc attagaataa atgtgtcctt 900

gatgtaaatg caaagcattt cttcctgatt aaattgtaga tgtagacttt acaatataat 960

tcaataataa aaagtaatta acctctaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaacac 1020

ttgttgggcg cggcgcgggc ccggagaaaa gtttttaaaa cccttcgtat ggcgcgagga 1080

ggggccccag gtaggggaac ggaacacagg ggccacggcg gtgtagcaag agggaaaacc 1140

cggtggcgaa gaaaaagggc gtggattctc acccaataaa aagcgcgccg ggggggaacc 1200

ggaaggcgtt tgggaaacaa gactctcaga aagagggaga ggcgtcagcg gaggcgcaca 1260

ataaaggggg gactcctcac aacaccgggg aggagctccc gaagaggaga cgccccgaaa 1320

tagacaacaa ttacagcccc cgggcgcgcc gggggcagat accagaagac gaagcacgag 1380

acgaagtagc acaagaagaa aggacagaca gaagcgagac gagaggacag aagggaggag 1440

agagaggcgg aggcgagggc gaaaacaagg ggacgacagc aacgagaaga gaaagaaaaa 1500

cacgacggag gaacggggcg gaaaaaggac gagaaagccg agagcaggcg ggcgggggca 1560

cacgggggaa acgaagagca aggaaagaag acaacaggag aggaggggag ggaagcgagc 1620

gaagagcgta gagacgccga gcaacaagaa gctagagaca gcagtagaca cggacagaca 1680

gacacggtga tggtagccgg ggcgggggga catcttgggc gcgaatgctt cgccggggaa 1740

cagacagcgc agggcgagca gagcgaggaa acgcagggaa cgacgcgaca gacaggaggt 1800

cagaagaagg gaagatgagt gcagcgggaa gaccaacgga gaaggagag 1849

<210> 157
<211> 903
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (139)..(139)
<223> a, c, g or t

<400> 157
ttgtttgttt tttttctttt tttacatttt ttttttttg tggacgaaac attcacacag 60

aaacagctag gaccagatta cgccagcttc atgagacctc tcctatctgg gcacgttgag 120

ttggctgact ctgggagcnc aggctgttgc ttcccagtct ggtggtgaat cctccatggt 180

ctggttgagt ccttttcaac tagtttttgt gttgtttttt gaacctcaag tcactgactc 240

cttcagcagt ggtttaagga agtatcttaa tttttaaatc acatgttgac catgcagcca 300

attgttggcc cgtagagtga acaaaaccag accacaagat tgagggcttg gagctggaga 360

aagggaaaag aaaaagcagg ctgtgactct ctggggaaag aactgaaaga tgacactagg 420

aattctcaaa gcgagaggaa aaggaaaggc ccttttcgg aaatgacctc tgataaacac 480

accctccagg gttacacctg ccactgtgtc ttccacggac agacaagctg cactttagca 540

gtcctgaata cctagagact tccttaacag agagtgggga atctcgtcat cttgcatggg 600

gatgggagct cgaagggaga acctcagcct tccagaaggt tatataaaac cagttgagaa 660

tttccctaag aatggagcag tggacaaaca attgttattg taatccaaat acatgagtct 720

acctacataa tggagaaatg ctaacttaca gaaagcgtca ggcttggtgt tctcgaactg 780

gttttaattt ccttttataa aattacagcg aattatacaa ttacattgat taattgtaat 840

cagttctgag tctagataac aaaacacaac acacacaaac aaacaaacaa cctctcctca 900

aac 903

<210> 158
<211> 368
<212> DNA
<213> Homo sapien

<400> 158
gcggcgcccg ggcaggtaca gtggcaacag cattagacta agtggaacat cccagcaggc 60

tgctttagaa tccgctcatt tgactagata cgatgtaatt ggctgtcttt aaaaaacgcg 120

cacacacaca caatctgata ggcatatctc atgcccattc aatatggaat gttcttcgct 180

tgctgaattt aagcctgtat tttaaggttt gtggttcctc ggccacaatg ggtgatgtca 240

ctgatagaac gaagctgagt ttccaagggt tggggctgtg caagagtaaa cactagagct 300

tgagttgtta tccagctggc aagcacggaa gtctttgaag aatgtaatgt aaaaagggaa 360

aagaatgt 368

<210> 159
<211> 1548
<212> DNA
<213> Homo sapien

<400> 159
gtcgcggccg aggtacatag acacaggaca attaataatt tggaaaacaa aagacttact 60

tatctccata tctgctaatt gtctcccaat ctccttaaaa tgcacttatt agcaatcatt 120

tttcaccacc catttaccca aacaacagga caaagactca atttcctatt ttatacaaaa 180

catgataagt cagccaagta ggttctggac cggacaacaa gggagatatt aaattatagt 240

atttatataa aagtggccca ttcgtgtgat acaaaacgtc tcattatgtg gaccaagaaa 300

catataaaat atatcaatat ataagttgga aaaaataaca aaaaagcaca cacattaata 360

aataatcata ttacacacac ccacatctat attctcttat atacacgcac acacaactgt 420

gtgtcttaac aacaaactct ccattttata aaatatctca ctgtatctct ttataaccac 480

aagaaagatg gaaacacaat aaacaccaaa aacaaagaca cacagaaatc atttgtgcgc 540

ctatatataa taatcacact cagtatattt aaccaacact tccccccca tatcatctaa 600

tttatgtagt actctataat gtatcatcta gtatattatg tagaaaagtg gcgccccata 660

tattatacag ccgggcgccc acgtgtatat caaacaacag ctgttgtcgt tctgcccggg 720

gagcacaaaa ggtgtataac aacaaatgat tatattcttc acccgaggaa gaagataatg 780

agaagggagt aataataacc agcgaagaag gggcgggggg acataataac accgctaatc 840

aacaccaaag aacgcgggcc aaacgaaaca acaacacaaa cgagatgttg tgccacacct 900

cagtaacacc tacatgtcag cgcgcagcga caaaacaagg gcatgaccag aagaaacctg 960

aaaaaactgc ggcaaaggga accacaccac cgacacacac accccaatga ggaaagagaa 1020

taaaaaacac atagttgagg acaccaaaaa aaggacagga ggaaaaacac actgtggggt 1080

gtgcaaaaca cacgttgtcg taacaacacc gcaggagatc acaaaagaaa acaccccaag 1140

acacaacaaa atataagcga acaccacaac aacacaacaa cgaaacaaaa caagaaaaaa 1200

caaaacacac aaccgaaaca aacaaccaaa acgaccacca gaagcacaca acacgagacc 1260

aggacaacac cagacgcaga cagacacgag atgatgcaca cgcgcatcgg ccgagccagc 1320

aaacggcacg ccaacaagca agacggccag cggagaacac gagcatgtcg aagtaggaca 1380

gcccgccaag acgacaaaaa accgaggaga agagaggaag aacacagaaa ccgcgaggaa 1440

aggacaacgc aggaggcgag ataacggcgg acaagcgcaa gaagggagca gagaagacag 1500

aaggcaagac acgaaacaaa aacacacaac gaacaaacac gcgagagg 1548

<210> 160
<211> 552
<212> DNA
<213> Homo sapien

<400> 160
agaagactga ctcatatagg gcgatggtca ctagatcatg ccgagcggcg ccattgtgat 60

ggatctagct ttggctactg ccggtagtgg acaatatggc acatggaaat taaaaagtcc 120

ataaacgtgc cctcctaaca cgagaataag aaaggtggct gaagtagata atttcagtga 180

cggaggggat gaaatatttt ttggttaatt gatgtaatga tgactcacta tgccttattt 240

cctatttta aaaacacaga atgagcaagt cattcctgaa caaaaattta ctgtgtgtat 300

aacatacacc tcaaaatgaa ttttaaggga acatattact aatcaaataa cacagtttat 360

gctttttcaa tttccacaaa ttgttaatta tgatacttaa gggaaccctt acaatatata 420

acaagtcatt tcaatattat tcatcatcct taacttctga aagtttggtt tatgttatct 480

tatctagaaa gaaaactact tacaaatctc attttcccac aaaattaatt caacatccaa 540

ccttaaaaat aa 552

<210> 161
<211> 3937
<212> DNA
<213> Homo sapien

<400> 161
tgtagctggg attacaggca tacaccacca cacccagcta ttttttttgta tttttgttag 60

agacaggttt ctccatgttg gtcaggctgg tctcgaactc ccgacctcag gtgatccgcc 120

cacctcggcc tcccaaagtg atgattttat ttttatttttt aattgtatta atctgcatct 180

agataatgac tttgtgaagg gtgtgtgttg tgtgcaaaac ttaaggtatt ggttgagagt 240

taaatacatt atttttataa tatgttgggt atagtctagt tactaatgat ttttttttaag 300

tacttttata aaaagttcat ttttaaaatt gtttgttttt aaaagccaat atacgttgca 360

gaattaggaa cagtatttat atttatttac acaagacatt gtgccatagc atcctagtaa 420

aacaccttca tgaatgagta atgttatctc ccagaattac attaaaatta tttctaaaaa 480

gtagcaaagt cattaccttt tgcttttaat gacccacacc tcaccagctc ctggtctttt 540

cttcactgtt gcccttattt tgaggcaatt tttcttaaaa tatgactttt atgcaccaca 600

tttagtagag gcagtgacat cagtgatctc agtacccatc agctgtcccc ctcctctgcc 660

cttcttcatc tcttctacct tgtgaccatt tcccttaccg gttcatgttc tcctttatct 720

ctgcttttct ttcttagcca ggatacttcc ctcacaactc tcactcccaa attcttttag 780

atatacattt ttctggatat tggctgctga aatctgaagc tctggtaaag ttcctagtat 840

cagagatcaa tcctggagga ggcctagttc actattagat tacaaagact cctcacaaag 900

taaaggaaaa tcaccttcaa aaccacaacc ctttatgttg tcaagtctaa tatgagtgtt 960

tttacgaagt atttctttct acccattgtt caagaatgta aatgtaaaaa aaaatacaag 1020

agagttgggt agatatgcat gcttgaggaa acttgctttt actgttttcc tacttgtatc 1080

cccagttcag ttgaatttac aaggacctac aagatggtca tgtttgtctt ggtatgtgct 1140

accccaattt tagtgtttct ttctttattt taaatcagta attattcagt tgattgttta 1200

tactatataa tgaagtaaca aaaacatttt ggtttgtatg ttttaagtaa cagttgtgca 1260

aattcctctt gtttgttagg tgctcccttt gaatattttg tgaactgtgt cagagggaga 1320

ggggtggtgg ctaggaagag ggtcagaaag aagctagagg gaggtcagga gaagggtaac 1380

agggaggatg caaagcagac atctaccctg gtcaccccag gatcaggata tctgtccttg 1440

tttcatgttg aattcaaaaa ttggatctca cttaggcttt gaaggtgaca gccatctctg 1500

atagctgagc ataagtaaag aaaggtggag tgccgatgca gaaaggaaaa tattcagctt 1560

tcctctctta gatcgcactt tgaagatggc cttttggaga caatctgaca ggttaaaaca 1620

ggaactgttg gaattattct agctgtaact acctattggc tatgtgttga ttgatcctag 1680

aaagaaaaaa taatttttca ttttagatct tgattgaatt taagatgtat ttatatgcct 1740

acaaaaggtc tgtcttgtaa ctgttgtata aaataaacct aatctatggt ttcattttta 1800

atctaaaaaa agttgtgcct taacaatagg gcattgtatg ttaataaggg aaaacaacct 1860

ttttagtaga tgggggaaaa taggaacttt ttgccattaa aacttaagtt cttttgatgt 1920

ttttaatatt atagttgggg gagattcatt aaaattaaat tgaaataaaa ttatttttgc 1980

ataacctagc atttacaact aaagtatgtt ttttataaga actggcatct tgatgtatat 2040

aggtctgaaa taatatttca tcttttgatt tttaatttta ataatattag accaggatag 2100

atcacagttt tacaaatctt agttttaaat aaattatttc agtgtgctgt tagtcctcta 2160

cagtcatttt ggtttaaaaa gtgactattt atttatggta gcatatcaat aatttattaa 2220

tgttaaaaaa tactgtgtat gacattacaa accagaacag ttcctggggg agaggattct 2280

aattgattgg cagttctgag agggcaagaa gaatggaact ttatacttca aaaggaggtt 2340

ttggttttac caggtactgc ttatgtaaat cgtttatttt tatttcatca aagcctggca 2400

agtatatgca ttccaattta ccattggcaa agctttattt atttttaagg ttggatgttg 2460

aattaatttt gtgggaaaat gagatttgta agtagttttc tttctagata agataacata 2520

aaccaaactt tcagaagtta aggatgatga ataatattga aatgacttgt tatatattgt 2580

aagggttccc ttaagtatca taattaacaa tttgtggaaa ttgaaaaagc ataaactgtg 2640

ttatttgatt agtaatatgt tcccttaaaa ttcattttga ggtgtatgtt atacacacag 2700

taaatttttg ttcaggaatg acttgctcat tctgtgtttt taaaaatagg aaataaggca 2760

tagtgagtca tcattacatc aattaaccaa aaaatatttc atcccctccg tcactgaaat 2820

tatctatttc agccaccttt cttattctcg tgttaggagg gcacgtttat ggacttttta 2880

atttccatgt gccatattgt ccactaccgg cagtagccaa agctagctgt ttcagtccca 2940

cagaagagac agtgctctgc catgatgaca gggcactgct agggctggtt tttcttgttt 3000

ttccctttg gcagtgtgga cttcaggaac tagatgtata tgcacaaggg attgagttta 3060

cactaaaact aggaaatgga gttttcaatc tatgttcttg cctcttcata cttttattta 3120

ttttttgtca tcctgcctta tactgggcta acaatgagat aaaataaaaa taccttttgaa 3180

tactcttttc cctttcatgc atttaaagcc atggaggaac tagaccatta gctgttgccg 3240

tcacatgctt agacaccagt ttacttagcg tgttatgacc ttcctcaccc atactaccaa 3300

atttaaatgg gtcccgactt caccctctgg aaggaagtaa actcttctct ccccatggtt 3360

tcagagcagt ttttacctgc aagcaccatc tctgtatgtg ctcttactag attatacagt 3420

tcttgagagg gattgcatct tggtgttttt gtatttccac ctcaccccca gcacatagcc 3480

cagtctcttg cacaaattaa gtacttaatg tgtgttgagc taaattgaat aaaggattat 3540

tagcattagc atattttgtg ccttggttgt ataagctggt tgtttgtttt gttacctttg 3600

caaatattta tgattatcac ccccccacat actaaattgt ttttaaaagt tttgcctttc 3660

cttcagatac taccccaggc aatttgctgt agataatgtg attgcttcca atgacataat 3720

tatcccaaac tctctgcccc ggatatactt tgccaaacga aatttgaatt ctctgaataa 3780

attggtcatg tctaaaaaaa aaaaaaaaaa aaaattgggg gcgggcgaaa aattttaaac 3840

atgtgggggg gggcaatgaa agaaggcagg tccaagacgg ggaaggggggt cacaaaaagc 3900

ggggggttga aaaggggggg tggtataggt aacgggg 3937

<210> 162
<211> 852
<212> DNA

<213> Homo sapien

<400> 162
cggccgcccg ggcaggtacc accatgccca gctaattttt atatttttag tagagatggg 60

gttttgccat attggccagg ctggtctcaa actcctgaac tcaagtgata acacccacct 120

tggcctccca aagtgctggg attacaggtg tgagccacca cactgggcca atgcttaata 180

ttttaatgta tctcaacaat aaaaccaaga agaaacaaag ccttttgact tgttagaatg 240

tattaagtag tattttaaag aaactttata gttgtgacat tgaaagactg ttggggtggg 300

gggaggaaaa tttttacttt ccatcttaat gtaaccttat gctattctgt atttttactg 360

tatattgctt ttacaataaa tataaaatga aaatgtttat gttgacaaaa agaacaaaaa 420

acaacaaaca acaaaaaaca aaaggctggg ggtgtcacac ctgtgggcca aaagctggtt 480

tccctggggg tggacatttg gttttatccc ggccccacaa ttccccaccc aaatattacc 540

gggagacaac gggaagaacg acacaacaca caaaaagaca caacacacaa aaccaccaca 600

cagcaacgcc cgctcatcgg aggcagcgca caagacgaga gcagaaggaa aagggacaac 660

aaaaaaaagc aagtagcacg atcacaacac agagccacga caagaagaga aggacaatga 720

cgaaaacgag cagcagcaca agacacagac aagacaaaag caagaaagac agaacacgac 780

agaagacaac aagagacgaa acaaccaaaa agccacacaa aaaagcgagc gaaaacaaga 840

accaactaac ac 852

<210> 163
<211> 685
<212> DNA
<213> Homo sapien

<400> 163
caccgatgat taaccagtac agacgctgtt ccactagatc atggcgagcg gttcgcagtg 60

ttatagtagt gcgcggctga catacttcat ctgcttttta ttgttgttgt tgtgtttttt 120

aaccaaaatg gcatttgttt ggctgtcatt gtcgagtata tatttattgt gttttacaaa 180

catgagtata tatgtatgta tatgtataaa ccaaacttat atatatagaa gtcaagggca 240

tgtatactag atattttaaa gagttattta tcaagggaaa aagatgtgtg ttataaagta 300

aacagagtct atattttcta tataatgtag atagtcaaac atagcttata aattatagga 360

ggttttggtt ttctttttta ttattaaagg aaaaaaagaa gaaaaaaaaa gacaatgaat 420

gcaactgttc ttagtgtttt taaggccaaa ctactgtgga agctgggagg cggccctccc 480

tgcgggcctc cagcagccct gtgcctgccg ctaggagctc cagagctgat gcccgttgtg 540

atctctgcga tgctcgatgc tcggagccag cggtcagcat cactcagcca gctggcctgt 600

gccgctctga cgtggcttcc agctgttctc cgcaatttgc attggtggga taaaggaatg 660

aaaaggataa ataaagattt aaaag 685

<210> 164
<211> 2396
<212> DNA
<213> Homo sapien

<400> 164
cggccactga attcccttgc ggccgcagaa ttttttttttt ttttttttgt attttctttt 60

aaatctttat ttatcctttt cattccttta tcccaccaat gcaaattgcg gagaacagct 120

ggaagccacg tcagagcggc acaggccagc tggctgagtg atgctgaccg ctggctccga 180

gcatcgagca tcgcagagat cacaacgggc atcagctctg gagctcctag cggcaggcac 240

agggctgctg gaggcccgca gggagggccg cctcccagct tccacagtag tttggcctta 300

aaaacactaa gaacagttgc attcattgtc ttttttttttc ttctttttttt cctttaataa 360

taaaaaagaa aaccaaaacc tcctataatt tataagctat gtttgactat ctacattata 420

tagaaaatat agactctgtt tactttataa cacacatctt tttcccttga taaataactc 480

tttaaaatat ctagtataca tgccttgact tcttatatat ataagtttgg ttttatacat 540

atacatacat atatactcat gtttgtaaaa cacaataaat atatactcga caatgacagc 600

caaacaaatg ccattttggt taaaaaacac aacaacaaca ataaaaagca gatgaagtac 660

tataaaagca ccaggcagca gacagaaagc cacatttgct agaaacttct ccctcccctg 720

ggcaaggcac agggcagaat actaaatacg tccaggtgcc tgaaagagaa ggaaggcagc 780

aaaggaatgg gtcagatcac aagcttttgt tttgtttctt agcctgggat tagaccaaga 840

atcacaagta agtcattgcg tttataagga aaaaccaagg ggctcattca acagcttagc 900

ccttgtggtg gctgcagggg acagtagaga cctggaaggg gaaggagaga gacagtgacc 960

tggtgacaac tccatgacta ttgtctagcc cctgcctaca tctgcccagg caagcttttg 1020

attgccacac taagcatcaa gccattgcat ataatatgtt cttttcatga ctctattcac 1080

tgctggtgta aacagaaagg agtattcagt agtgcaacac tttgggaggc tgaatgagag 1140

ccacagtttc ggtggccccc tcaagggctc cagcaaatgg ggactgtgac gtgggggaga 1200

aaaaggccag tgggccattg ccttccactc ctgagacagt accggcttca cctccagatg 1260

ccactgggaa cactgagccc atcacctttt aaagaagcgc aggaggtctt gactcttcct 1320

atgagaagtc ccagaagccc cagtcagtct gggtgggggg tcccctgaaa ctcagagcac 1380

gctccagctg ctggctcctg tgcctggcgg ggagaaggtg gagaggcggg catgcctgcc 1440

gctgagaaag acataataac gctccagctg cacacccctt ttcttccacg gccctcctct 1500

ttctccccaa aaggaaacta aaatgtgggg ttctgatcat tgatttttaa acaagctccc 1560

caatgtcatg gcctctgctc tgtaaactcc aaggctctga tggccaaaaa cgtaggcggt 1620

ggtggagcag taggcggtgt caaggctaga tttgtattaa tatctaggag tggggtgggg 1680

atgcaggtgg agggtgagcc aagtagacca agcctcagga aatcccagag gtgaagattt 1740

gatgcacatc ccccatcctt gccaaataaa atacaagtta gaggtcattg ttcacattac 1800

agtcacaatg ggacatttta aacagctcgt tttctgagta cattcaagac ataattgaat 1860

ttattttaaa aaatggattt cccggctggg cacggtggat cacgcctgtc atcccagcac 1920

tttgggaggt caaggaagcc ccttcttcac aggacatcct ggtgttcctc actgggcagg 1980

aggagatcga agccatgagc aagacgtgcc gagacattgc aaagcacctc ccagacggct 2040

gccctgcgat gctggtcctt cctctgtacg cctccctgcc ctatgcacag cagctccgag 2100

tcttccaagg ggccccaaag ggctatcgca aagtgatcat ttcaaccaac atcgctgaaa 2160

cctccataac cattacagga ataaaatatg tagttgacac gggcatggtt aaagcaaaga 2220

agtataaccc tgacagtggt cttgaggtgt tagcagtgca gcgggtatcg aagacgcagg 2280

cttggcagcg cacagggagg gctggcagag aggacagtgg catctgctac cggctctaca 2340

cggaggacga gtttgagaag tttgataaga tgaccgtgcc agagatccag aggtag 2396

<210> 165
<211> 11
<212> PRT
<213> Homo sapien

<400> 165

Met Arg Tyr Leu Pro Gly Leu Ser Ala Arg Ile
1 5 10

<210> 166
<211> 45
<212> PRT
<213> Homo sapien

<400> 166

Met Ser Ile Pro Arg Ala Glu Ile Ser Leu Leu Glu Ser Phe Gln Leu
1 5 10 15

Thr Ser Thr Val Ala Thr Ser Glu Ser His Lys Ser Asn Gly Ser Cys
20 25 30

Arg Lys Pro His Leu Leu His Cys Pro Arg Ile Asn Gln
35 40 45

<210> 167
<211> 37
<212> PRT
<213> Homo sapien

<400> 167

Met Ile Leu Gly Ser Asp Asn Gly Ile Arg Arg Ile Lys Tyr Leu Gly
1 5 10 15

Ile Gln Tyr Tyr Ala Cys Ser Phe Phe Gln Ile Val His Gly Gly Gly
20 25 30

Gly Cys Val Ser Gly
35

<210> 168
<211> 82
<212> PRT
<213> Homo sapien

<400> 168

Ser Leu Ser Val Ala Gln Ala Arg Val Gln Trp Arg Asp Pro Gly Ser
1 5 10 15

Leu Gln Pro Leu Pro Pro Gly Phe Lys Arg Phe Leu Ser Leu Ser Leu
20 25 30

Pro Ser Ser Ala Gly Tyr Arg Arg Ala Pro Pro Pro Cys Pro Ala Leu
35 40 45

Leu Tyr Phe Ala Val Glu Thr Gly Phe His His Val Gly Gln Ala Gly
50 55 60

Leu Glu Leu Leu Thr Ser Gly Asn Pro Ala Pro Pro Arg Pro Pro Lys
65 70 75 80

Val Leu

<210> 169
<211> 103
<212> PRT
<213> Homo sapien

<400> 169

Met Ala Ile Phe Ser Ala Leu Ser Gln Leu Leu Glu His Gly Leu Asp
1 5 10 15

Leu Glu Thr Ser Asn Lys Asp Phe Thr Ser Ile Pro Ala Ala Cys Trp
20 25 30

Trp Val Ile Ile Ser Met Thr Thr Val Gly Tyr Gly Asp Met Tyr Pro
35 40 45

Ile Thr Val Pro Gly Arg Ile Leu Gly Gly Val Cys Val Val Ser Gly
50 55 60

Ile Val Leu Leu Ala Leu Pro Ile Thr Phe Ile Tyr His Ser Phe Val
65 70 75 80

Gln Cys Tyr His Glu Leu Lys Phe Arg Ser Ala Arg Tyr Ser Arg Ser
85 90 95

Leu Ser Thr Glu Phe Leu Asn
100

<210> 170
<211> 131
<212> PRT
<213> Homo sapien

<400> 170

Arg Thr Ala Arg His Asp Tyr Ala Ser Cys Arg His Leu Met Val Phe
1 5 10 15

Ser Thr Arg Leu Thr Leu Lys Arg Cys Tyr Arg Glu Met Val Met Leu
20 25 30

Leu Val Phe Ile Cys Val Ala Met Ala Ile Phe Ser Ala Leu Ser Gln
35 40 45

Leu Leu Glu His Gly Leu Asp Leu Glu Thr Ser Asn Lys Asp Phe Thr
50 55 60

Ser Ile Pro Ala Ala Leu Leu Trp Val Ile Ile Ser Met Thr Thr Val
65 70 75 80

Gly Tyr Gly Asp Met Tyr Pro Ile Thr Val Pro Gly Arg Ile Leu Gly

85 90 95

Gly Val Cys Val Val Ser Gly Ile Val Leu Leu Ala Leu Pro Ile Thr
100 105 110

Phe Ile Tyr His Ser Phe Val Gln Cys Tyr His Glu Leu Lys Phe Arg
115 120 125

Ser Ala Arg
130

<210> 171
<211> 23
<212> PRT
<213> Homo sapien

<400> 171

Met Val Ala His Cys Ser Leu Pro Val Pro Val Thr Leu Pro Asn Cys
1 5 10 15

Pro Val Ala Cys Arg Leu Ala
20

<210> 172
<211> 57
<212> PRT
<213> Homo sapien

<400> 172

Met Asn Gly Arg Gly Leu Ala Arg Gln Gly Cys Glu Ser Gly Asn Ala
1 5 10 15

Phe Phe Thr Pro Met Asn Phe Cys Leu Ile Leu Thr Thr Glu Gln Glu
20 25 30

Cys Ser Glu Thr Glu Val Gly Val Thr Asn Ile His Phe Pro Phe Ser
35 40 45

Ser Cys Ser Asn Gln Tyr Phe Lys Lys
50 55

<210> 173
<211> 50
<212> PRT
<213> Homo sapien

<400> 173

```
Met Ser Gly Glu Arg Thr Asp Tyr Ala Pro Leu Asn Asn Lys Leu Cys
1               5                   10                  15

Ser Arg Val Arg Ala Val Ser Ser Glu Ala Leu Leu Thr Glu Thr Ala
            20                  25                  30

Thr Phe Lys Pro Asn Gln Arg Lys Tyr Leu Trp Asn Ser Leu Pro Gln
        35                  40                  45

Phe Gly
    50
```

<210> 174
<211> 31
<212> PRT
<213> Homo sapien

<400> 174

```
Met Val Thr Asn Leu Leu Thr Ala Gln Cys Gly Trp Met Val Ala Ala
1               5                   10                  15

Arg Cys Phe Cys Cys Leu Asn Leu Leu Asn Cys Leu Ile Phe Ser
            20                  25                  30
```

<210> 175
<211> 456
<212> PRT
<213> Homo sapien

<400> 175

```
Met Glu Pro Arg Cys Pro Pro Pro Cys Gly Cys Cys Glu Arg Leu Val
1               5                   10                  15

Leu Asn Val Ala Gly Leu Arg Phe Glu Thr Arg Ala Arg Thr Leu Gly
            20                  25                  30

Arg Phe Pro Asp Thr Leu Leu Gly Asp Pro Ala Arg Arg Gly Arg Phe
        35                  40                  45

Tyr Asp Asp Ala Arg Arg Glu Tyr Phe Phe Asp Arg His Arg Pro Ser
    50                  55                  60

Phe Asp Ala Val Leu Tyr Tyr Tyr Gln Ser Gly Gly Arg Leu Arg Arg
65                  70                  75                  80
```

```
Pro Ala His Val Pro Leu Asp Val Phe Leu Glu Glu Val Ala Phe Tyr
                85                  90                  95

Gly Leu Gly Ala Ala Ala Leu Ala Arg Leu Arg Glu Asp Glu Gly Cys
            100                 105                 110

Pro Val Pro Pro Glu Arg Pro Leu Pro Arg Arg Ala Phe Ala Arg Gln
        115                 120                 125

Leu Trp Leu Leu Phe Glu Phe Pro Glu Ser Ser Gln Ala Ala Arg Val
    130                 135                 140

Leu Ala Val Val Ser Val Leu Val Ile Leu Val Ser Ile Val Val Phe
145                 150                 155                 160

Cys Leu Glu Thr Leu Pro Asp Phe Arg Asp Asp Arg Asp Gly Thr Gly
                165                 170                 175

Leu Ala Ala Ala Ala Ala Ala Gly Pro Phe Pro Ala Pro Leu Asn Gly
            180                 185                 190

Ser Ser Gln Met Pro Gly Asn Pro Pro Arg Leu Pro Phe Asn Asp Pro
        195                 200                 205

Phe Phe Val Val Glu Thr Leu Cys Ile Cys Trp Phe Ser Phe Glu Leu
    210                 215                 220

Leu Val Arg Leu Leu Val Cys Pro Ser Lys Ala Ile Phe Phe Lys Asn
225                 230                 235                 240

Val Met Asn Leu Ile Asp Phe Val Ala Ile Leu Pro Tyr Phe Val Ala
                245                 250                 255

Leu Gly Thr Glu Leu Ala Arg Gln Arg Gly Val Gly Gln Gln Ala Met
            260                 265                 270

Ser Leu Ala Ile Leu Arg Val Ile Arg Leu Val Arg Val Phe Arg Ile
        275                 280                 285

Phe Lys Leu Ser Arg His Ser Lys Gly Leu Gln Ile Leu Gly Gln Thr
    290                 295                 300

Leu Arg Ala Ser Met Arg Glu Leu Gly Leu Leu Ile Phe Phe Leu Phe
305                 310                 315                 320
```

Ile Gly Val Val Leu Phe Ser Ser Ala Val Tyr Phe Ala Glu Val Asp
325 330 335

Arg Val Asp Ser His Phe Thr Ser Ile Pro Glu Ser Phe Trp Trp Ala
340 345 350

Val Val Thr Met Thr Thr Val Gly Tyr Gly Asp Met Ala Pro Val Thr
355 360 365

Val Gly Gly Lys Ile Val Gly Ser Leu Cys Ala Ile Ala Gly Val Leu
370 375 380

Thr Ile Ser Leu Pro Val Pro Val Ile Val Ser Asn Phe Ser Tyr Phe
385 390 395 400

Tyr His Arg Glu Thr Glu Gly Glu Glu Ala Gly Met Phe Ser His Val
405 410 415

Asp Met Gln Pro Cys Gly Pro Leu Glu Gly Lys Ala Asn Gly Gly Leu
420 425 430

Val Asp Gly Glu Val Pro Glu Leu Pro Pro Pro Leu Trp Ala Pro Pro
435 440 445

Gly Lys His Leu Val Thr Glu Val
450 455

<210> 176
<211> 28
<212> PRT
<213> Homo sapien

<400> 176

Met Ser Tyr Asn Ser Lys Leu Glu Ser Ile Arg Leu Lys Arg Val Ser
1 5 10 15

Met Lys Thr Ile Pro Lys Ile Pro Phe Thr Gln Asn
20 25

<210> 177
<211> 91
<212> PRT
<213> Homo sapien

<400> 177

```
Met Ala Leu Gly Ser Met Tyr Leu Val Leu Thr Leu Ile Val Ala Glu
1               5                  10                  15

Val Leu Arg Gly Ala Glu Pro Cys Cys Gly Pro Leu Lys Tyr Arg Val
            20                  25                  30

Leu Arg Pro Cys Pro Leu Pro Val His Cys Ala Pro Pro His His Gln
        35                  40                  45

Pro Ser Arg Gly Asn Pro Val Ala Cys Leu Pro Thr Tyr Lys Val Val
    50                  55                  60

Tyr Gln Ala Ala Val Leu Ala Thr Ala Phe Lys Phe Gln Cys Asp Leu
65                  70                  75                  80

Pro Gly Arg Ser Ile Thr Leu Arg Arg Ser Ala
                85                  90
```

<210> 178
<211> 54
<212> PRT
<213> Homo sapien

<400> 178

```
Met Lys Phe Ser Ser Ala Phe Val Gln Ser Lys Pro Leu Ser Ser Cys
1               5                  10                  15

Arg Ala Glu Thr Leu Tyr Met Lys Thr Val Ser Glu Leu Val Leu Ala
            20                  25                  30

Ser Ile His Glu Asn Cys Leu Ser Cys Met Leu Ala Lys Thr Ser Ser
        35                  40                  45

Glu Thr Lys Lys Leu Lys
    50
```

<210> 179
<211> 88
<212> PRT
<213> Homo sapien

<400> 179

```
Gly Arg Val Arg Phe Val Val Glu Leu Ala Asp Pro Lys Leu Glu Val
1               5                  10                  15

Lys Trp Tyr Lys Asn Gly Gln Glu Ile Arg Pro Ser Thr Lys Tyr Ile
```

20 25 30

Phe Glu His Lys Gly Cys Gln Arg Ile Leu Phe Ile Asn Asn Cys Gln
35 40 45

Met Thr Asp Asp Ser Glu Tyr Tyr Val Thr Ala Gly Asp Ala Lys Cys
50 55 60

Ser Thr Glu Leu Phe Val Arg Glu Pro Pro Phe Met Val Pro Ser Ser
65 70 75 80

Trp Ile Glu Thr Pro Ala Asp Cys
85

<210> 180
<211> 26
<212> PRT
<213> Homo sapien

<400> 180

Met Val Leu Tyr Ser Glu Gly His Gln His Gly Pro His Leu Leu Asn
1 5 10 15

Met Glu Asn Gln Asn Leu Asn Glu Tyr Asn
20 25

<210> 181
<211> 57
<212> PRT
<213> Homo sapien

<400> 181

Met Thr Glu Arg Ala Asp Gly Lys Ser Gln Ser Cys Ile Glu Glu Ile
1 5 10 15

Ser Met Val Ala Leu Lys Leu Leu Lys Pro Asp Val Ser Ser Ala Ser
20 25 30

His Trp Lys Met Asp Arg Trp Ala Asn His His Leu Thr Ser Gln Arg
35 40 45

Glu Gly Gln Cys Ala Lys Val Phe Lys
50 55

<210> 182
<211> 67

<212> PRT
<213> Homo sapien

<400> 182

Met Asn Thr Lys Ala Leu Pro Thr Thr Pro Ala Gln Thr Ala Ile Ser
1 5 10 15

Pro Pro Glu Gly Gln Cys Ser Ser Ser Ile Gly Leu Glu Thr Ile Pro
20 25 30

Glu Ser Pro Cys Phe Arg Thr Pro Glu Ser Ser Asn Ser Pro Ser Leu
35 40 45

Arg Arg Asp Leu Leu Ala Ala Lys Arg Val Lys Leu Ile Val Leu Gln
50 55 60

Ser Ser Ala
65

<210> 183
<211> 91
<212> PRT
<213> Homo sapien

<400> 183

Asp Val Gly Gly Ala Gln Val Leu Ala Thr Gly Lys Thr Pro Gly Ala
1 5 10 15

Glu Ile Asp Phe Lys Tyr Ala Leu Ile Gly Thr Ala Val Gly Val Ala
20 25 30

Ile Ser Ala Gly Phe Leu Ala Leu Lys Ile Cys Met Ile Arg Arg His
35 40 45

Leu Phe Asp Asp Asp Ser Ser Asp Leu Lys Ser Thr Pro Gly Gly Leu
50 55 60

Ser Asp Thr Ile Pro Leu Lys Lys Arg Ala Pro Arg Arg Asn His Asn
65 70 75 80

Phe Ser Lys Arg Asp Ala Gln Val Ile Glu Leu
85 90

<210> 184
<211> 101
<212> PRT

<213> Homo sapien

<400> 184

Met Arg Pro Gly Arg Tyr Gln Ala Pro Arg Pro Phe Leu Tyr His Gly
1 5 10 15

Cys Trp Val Thr Ser Gly Ser His His Leu Phe Pro Ser Leu Phe Pro
20 25 30

Ile Ser Gln Met Trp Gly His Gly Leu Asp Asp Gly Leu His Arg Ser
35 40 45

Phe His Leu Cys Glu Ser Lys Ser Gly Gln Ser Ala Arg Thr His Leu
50 55 60

Cys Pro Gly Ser Ala Pro Gln Asn Gln Pro Pro Ala Ser Leu Lys Gln
65 70 75 80

Lys Pro His Leu Lys Gly Cys Ser Glu Glu Ser Thr Phe Ser Met Ser
85 90 95

Cys Cys Trp Lys Ile
100

<210> 185
<211> 489
<212> PRT
<213> Homo sapien

<400> 185

Gly Trp Thr Val Ile Gln Asn Arg Gln Asp Gly Ser Val Asp Phe Gly
1 5 10 15

Arg Lys Trp Asp Pro Tyr Lys Gln Gly Phe Gly Asn Val Ala Thr Asn
20 25 30

Thr Asp Gly Lys Asn Tyr Cys Gly Leu Pro Gly Asn Glu Gln Ala Cys
35 40 45

Lys Ile Lys Ser Phe Tyr Leu Lys Trp Asp Phe Phe Ala Leu Lys Asn
50 55 60

Ile His Cys Trp Lys Pro Val Leu Gly Ser Ala Glu Glu Phe Pro Asp
65 70 75 80

```
Lys Asn Val Glu Ala Lys Asp Lys Gly Arg Lys Ala Val Phe Ser Phe
                85                  90                  95

Pro Lys Phe Tyr Phe Trp Ala Glu Ile Leu Phe Cys Phe Ser Phe Gly
            100                 105                 110

Glu Tyr Trp Leu Gly Asn Asp Lys Ile Ser Gln Leu Thr Arg Met Gly
        115                 120                 125

Pro Thr Glu Leu Leu Ile Glu Met Glu Asp Trp Lys Gly Asp Lys Val
    130                 135                 140

Lys Ala His Tyr Gly Gly Phe Thr Val Gln Asn Glu Ala Asn Lys Tyr
145                 150                 155                 160

Gln Ile Ser Val Asn Lys Tyr Arg Gly Thr Ala Gly Asn Ala Leu Met
                165                 170                 175

Asp Gly Ala Ser Gln Leu Met Gly Glu Asn Arg Thr Met Thr Ile His
            180                 185                 190

Asn Gly Met Phe Phe Ser Thr Tyr Asp Arg Asp Asn Asp Gly Trp Tyr
        195                 200                 205

Val Trp His Ser Leu Leu Leu Leu Ala Lys Ser His Ala Tyr His Tyr
    210                 215                 220

Ser Glu Ser Leu Thr Ile Phe Leu Ile Ala Thr Thr Ser Trp Ala Leu
225                 230                 235                 240

Thr Val Ser His Cys Pro Lys Leu Phe Met His His Ser Lys Ala Phe
                245                 250                 255

Gln Leu Ala Gly Arg His Ser Tyr Ser His Phe Thr Asp Glu Ile Ala
            260                 265                 270

Arg Asp Tyr Val Ile Cys Pro Met Ser His Asn Tyr Pro Glu Ile Lys
        275                 280                 285

Leu Glu Phe Glu His Ser Tyr Phe Leu Asn Asn Glu His Leu Asp Lys
    290                 295                 300

Tyr Leu Tyr Leu Tyr Ile Leu Lys Cys Val Ala Lys Leu Ser Phe Ser
305                 310                 315                 320
```

```
Phe Pro Gly Phe Ser Asp Thr Lys Gly Cys Lys Ser Tyr Tyr Ser Ser
                325                 330                 335

Ile Lys Ala Gln Thr Gln Ser Leu Asp Gly Leu Pro Gln Arg Pro Ser
            340                 345                 350

Tyr Leu Ser Phe Leu Leu Ala Gly Thr Gly Gly Leu Trp Cys Ile Ser
        355                 360                 365

Val Thr Leu Cys Ile Ala Pro Lys Gly Lys Thr Thr Val His Thr Ser
    370                 375                 380

Val Ala Val Phe Tyr Gly Ala Ser Ala Lys Arg Asn Leu Thr Thr Val
385                 390                 395                 400

Val Leu Phe Leu Ile Thr Pro Asn Thr Phe Ser Phe Arg Leu Thr Ser
                405                 410                 415

Asp Pro Arg Lys Gln Cys Ser Lys Glu Asp Gly Gly Gly Trp Trp Tyr
            420                 425                 430

Asn Arg Cys His Ala Ala Asn Pro Asn Gly Arg Tyr Tyr Trp Gly Gly
        435                 440                 445

Gln Tyr Thr Trp Asp Met Ala Lys His Gly Thr Asp Asp Gly Val Val
    450                 455                 460

Trp Met Asn Trp Lys Gly Ser Trp Tyr Ser Met Arg Lys Met Ser Met
465                 470                 475                 480

Lys Ile Arg Pro Phe Phe Pro Gln Gln
                485
```

<210> 186
<211> 33
<212> PRT
<213> Homo sapien

<400> 186

```
Met Val Thr Glu Ser Leu Ser Ser Pro His Ser Glu Ser Ile Pro Leu
1               5                   10                  15

Gly Arg Val Asn Pro Gly Ser Gly Leu Pro Pro His Ser Thr Arg Pro
            20                  25                  30
```

Phe

<210> 187
<211> 149
<212> PRT
<213> Homo sapien

<400> 187

Pro Gly Asn Leu Asp Thr Ser Ser Arg Gly Ser Ser Gly Ser Pro Ala
1 5 10 15

His Ala Glu Ser Tyr Ser Ser Gly Gly Gly Gly Gln Gln Lys Phe Arg
20 25 30

Val Asp Met Pro Gly Ser Gly Ser Ala Phe Ile Pro Thr Ile Asn Ala
35 40 45

Ile Thr Thr Ser Gln Asp Leu Gln Trp Met Val Gln Pro Thr Val Ile
50 55 60

Thr Ser Met Ser Asn Pro Tyr Pro Arg Ser His Pro Tyr Ser Pro Leu
65 70 75 80

Pro Gly Leu Ala Ser Val Ala Gly His Met Ala Leu Pro Arg Pro Gly
85 90 95

Val Ile Lys Thr Ile Gly Thr Thr Val Gly Arg Arg Arg Arg Asp Glu
100 105 110

Gln Leu Ser Pro Glu Glu Glu Glu Lys Arg Arg Ile Arg Arg Glu Arg
115 120 125

Asn Lys Leu Ala Ala Ala Lys Cys Arg Asn Arg Arg Arg Glu Leu Thr
130 135 140

Glu Lys Leu Gln Ala
145

<210> 188
<211> 41
<212> PRT
<213> Homo sapien

<400> 188

Met Thr Val Pro Leu His Thr Ser Leu Ser Tyr Arg Gly Arg Ser Gln
1 5 10 15

Leu Leu Lys Thr Lys Thr Thr Ile Asn Ile Tyr Lys Asn His Asn Ile
20 25 30

Lys Gly Phe Met Leu Arg Lys Asn Pro
35 40

<210> 189
<211> 45
<212> PRT
<213> Homo sapien

<400> 189

Met Tyr Thr Asn Lys Tyr Ala Gln Asp Leu Glu Ser Tyr Ile Lys Met
1 5 10 15

Tyr Leu Thr Trp Leu Glu Cys Val Cys Val Phe Pro Arg Leu Ser Lys
20 25 30

Ile Arg Lys Pro Glu Ser Gln Ala Thr Lys Lys Lys Asn
35 40 45

<210> 190
<211> 91
<212> PRT
<213> Homo sapien

<400> 190

Met Phe Leu Cys Asn Val Leu Arg Val Thr Trp Ala Ser Pro Thr Tyr
1 5 10 15

Ala Ser Thr Val Cys Cys Val Thr Phe Arg Gln Leu His Thr Pro Pro
20 25 30

Ala Pro Leu Pro Ser Pro Pro Ser Ser His Thr Val Ser Ala Gly Cys
35 40 45

Gly Ser Pro Thr Ser Val Met Ser Gly Ile Met Leu Leu Leu Ser Leu
50 55 60

Leu Phe Ser Leu Phe Phe Phe Phe Val Ile Gln Val Leu Leu Thr Ser
65 70 75 80

Ser Leu Ile His Gln Asn Ala Arg Ser Ser Tyr

85 90

<210> 191
<211> 100
<212> PRT
<213> Homo sapien

<400> 191

Ala Asp Asn Asp Ile Gly Ala Val Ser Thr Thr Gly His Gly Glu Ser
1 5 10 15

Ile Leu Lys Val Asn Leu Ala Arg Leu Thr Leu Phe His Ile Glu Gln
20 25 30

Gly Lys Thr Val Glu Glu Ala Ala Asp Leu Ser Leu Gly Tyr Met Lys
35 40 45

Ser Arg Val Lys Gly Leu Gly Gly Leu Ile Val Val Ser Lys Thr Gly
50 55 60

Asp Trp Val Ala Lys Trp Thr Ser Thr Ser Met Pro Trp Ala Ala Ala
65 70 75 80

Lys Asp Gly Lys Leu His Phe Gly Ile Asp Pro Asp Asp Thr Thr Ile
85 90 95

Thr Asp Leu Pro
100

<210> 192
<211> 54
<212> PRT
<213> Homo sapien

<400> 192

Met Glu Glu Gln Glu Glu Ala Leu Cys Ser His His Ile Pro Val Ala
1 5 10 15

Arg Ser Trp Leu Gln Gly Ser Ser Gly Asn Arg Ile Pro Arg Ser His
20 25 30

Glu Thr Ser Pro Asn Ser Ala Val Thr Glu Ser Thr Arg Gln Trp Leu
35 40 45

Lys Asp Gly Glu Thr Ser
50

<210> 193
<211> 63
<212> PRT
<213> Homo sapien

<400> 193

Met Ile Ile Leu Lys Tyr Arg Trp Lys Asp Thr Asn Ala Arg Lys Arg
1 5 10 15

Glu Ser Asn Gln Pro Arg Phe Gly Gly Trp Gly Thr Glu Asp Gly Ala
20 25 30

Thr Phe Pro Pro Tyr Leu Leu Phe Phe Tyr Ile Pro Ile Cys Thr Leu
35 40 45

Arg Ile His Leu Arg Ser Ser Phe Lys Arg Glu Lys Leu Asp Thr
50 55 60

<210> 194
<211> 211
<212> PRT
<213> Homo sapien

<400> 194

Met Val Phe Leu Lys Phe Phe Cys Met Ser Phe Phe Cys His Leu Cys
1 5 10 15

Gln Gly Tyr Phe Asp Gly Pro Leu Tyr Pro Glu Met Ser Asn Gly Thr
20 25 30

Leu His His Tyr Phe Val Pro Asp Gly Asp Tyr Glu Glu Asn Asp Asp
35 40 45

Pro Glu Lys Cys Gln Leu Leu Phe Arg Val Ser Asp His Arg Arg Cys
50 55 60

Ser Gln Gly Glu Gly Ser Gln Val Gly Ser Leu Leu Ser Leu Thr Leu
65 70 75 80

Arg Glu Glu Phe Thr Val Leu Gly Arg Gln Val Glu Asp Ala Gly Arg
85 90 95

Val Leu Glu Gly Ile Ser Lys Ser Ile Ser Tyr Asp Leu Asp Gly Glu
100 105 110

Glu Ser Tyr Gly Lys Tyr Leu Arg Arg Glu Ser His Gln Ile Gly Asp
115 120 125

Ala Tyr Ser Asn Ser Asp Lys Ser Leu Thr Glu Leu Glu Ser Lys Phe
130 135 140

Lys Gln Gly Gln Glu Gln Asp Ser Arg Gln Glu Ser Arg Leu Asn Glu
145 150 155 160

Asp Phe Leu Gly Met Leu Val His Thr Arg Ser Leu Leu Lys Glu Thr
165 170 175

Leu Asp Ile Ser Val Gly Leu Arg Asp Lys Tyr Glu Leu Leu Ala Leu
180 185 190

Thr Ile Arg Ser His Gly Thr Arg Leu Gly Arg Leu Lys Asn Asp Tyr
195 200 205

Leu Lys Val
210

<210> 195
<211> 54
<212> PRT
<213> Homo sapien

<400> 195

Met Asp Asp Ser Lys Leu Gln Lys Lys Lys Asp Val Asp Lys His Cys
1 5 10 15

Leu Thr Glu His Phe Ile Phe Ser Gln Leu Phe Trp Phe Leu Leu Ile
20 25 30

Thr Met Thr Lys Met Leu Asp Ser Glu Leu Cys Arg Tyr Phe Ser Lys
35 40 45

Phe Tyr Asp Phe Lys Ser
50

<210> 196
<211> 88
<212> PRT
<213> Homo sapien

<400> 196

```
Met Leu Gly Leu Gln Thr Leu Ser Arg Phe Leu Ser Gly His Pro Gly
1               5                  10                  15

Phe Leu Thr His Cys Leu Lys Ser Arg Trp Gln Val Pro Ser Leu Asn
            20                  25                  30

His Ser Cys Ala Pro Glu Asp Ser Gly Pro Lys Leu Pro Ser Ser Ala
        35                  40                  45

Cys His Ser Leu Leu Ile Ile Ser Ser Ser Asp Gln Val Cys Val Met
    50                  55                  60

His Leu Ala Gln Ala Gln Gly Val Pro Arg Arg Asp His Asp Pro Ser
65                  70                  75                  80

His Cys Ala Arg Ser Ser Ser Ile
                85
```

<210> 197
<211> 48
<212> PRT
<213> Homo sapien

<400> 197

```
Met Thr Glu Met Thr Gln Ser Lys Gly Arg Ile Gly Thr Glu Asp Ala
1               5                  10                  15

Asn Thr Gly Ser Tyr Lys Ile Gln Arg Glu Leu Ser Gly Gly Lys Thr
            20                  25                  30

Gln Glu Pro Asn Ser Thr His Leu Ile Pro Leu Val Asp Gln Leu Asn
        35                  40                  45
```

<210> 198
<211> 121
<212> PRT
<213> Homo sapien

<400> 198

```
Phe Phe Ala Asp Glu Val Ser Arg Leu Ser Pro Gly Leu Glu Cys Ser
1               5                  10                  15

Gly Val Ile Ser Ala His Cys Asn Phe His Leu Leu Gly Ser Ser Ser
            20                  25                  30

Ser Pro Ala Ser Ala Ser Gln Val Ala Glu Ile Thr Gly Ala Cys His
```

35 40 45

Pro Thr Trp Leu Ile Phe Val Ile Leu Val Glu Thr Gly Phe His His
50 55 60

Val Gly Gln Ala Asp Ala Leu Leu Thr Ser Gly Asp Pro Pro Phe Ser
65 70 75 80

Ala Ser Gln Ser Ala Gly Ile Thr Gly Val Ser His Arg Ala Arg Pro
85 90 95

Ala Asn Thr Phe Ala Leu Thr Thr Leu Gly Leu Leu Tyr Lys Ile Val
100 105 110

Met Ile Ala Met Glu Val Leu Pro Pro
115 120

<210> 199
<211> 162
<212> PRT
<213> Homo sapien

<400> 199

Met Asp Ala Ala Gly Gln Val Leu Gly Pro Glu Arg Gly Gly Tyr Leu
1 5 10 15

Pro His Trp Val Ala Ser Ser Ala Ala Pro His Leu Ser Leu Phe Ser
20 25 30

Pro Lys Leu Val Phe Leu Thr Ile Ile Val Val Gly Gly Gly Gln Met
35 40 45

Leu Lys Val Glu Ala Asp Leu Glu Lys Glu Thr His Gly Val Thr Val
50 55 60

Ala Lys Asp Ser Trp Lys Arg Asn Ser Ile Thr Ser Ser Leu Ala Thr
65 70 75 80

Thr Arg His Pro Arg Pro Trp His Ser Gln Arg Leu Cys Ala Val Ala
85 90 95

Lys Pro Leu Asn Leu Phe Trp Pro Cys Val Leu Gln Arg Ser Leu Cys
100 105 110

Cys Lys Thr Val Asp Ser Phe Asp Glu Val Leu Lys Asn Ala Thr Arg

115 120 125

Gly Gly Gly Val Trp Leu Ala Val Trp Pro Ser Ser Glu Lys Val Ala
130 135 140

Glu Ile Arg Gly Gln Gly Cys His Ser Pro Arg Leu Ser Ser Gly Ser
145 150 155 160

Gln Ser

<210> 200
<211> 594
<212> PRT
<213> Homo sapien

<400> 200

Val Pro Gly Arg Lys Leu His Arg Ser Arg Leu Gln Ala Ala Ala Pro
1 5 10 15

Arg Pro Ser Thr Cys Ala Gln Ser Leu Cys Trp Ser Arg Pro Pro Ala
20 25 30

Ala Gly Thr Gly Thr Gly Asp Pro Ser Gln Ser Lys Ala Pro Thr Met
35 40 45

Ala Met Gly Leu Phe Arg Val Cys Leu Val Val Val Thr Ala Ile Ile
50 55 60

Asn His Pro Leu Leu Phe Pro Arg Glu Asn Ala Thr Val Pro Glu Asn
65 70 75 80

Glu Glu Glu Ile Ile Arg Lys Met Gln Ala His Gln Glu Lys Leu Gln
85 90 95

Leu Glu Gln Leu Arg Leu Glu Glu Glu Val Ala Arg Leu Ala Ala Glu
100 105 110

Lys Glu Ala Leu Glu Gln Val Ala Glu Glu Gly Arg Gln Gln Asn Glu
115 120 125

Thr Arg Val Ala Trp Asp Leu Trp Ser Thr Leu Cys Met Ile Leu Phe
130 135 140

Leu Met Ile Glu Val Trp Arg Gln Asp His Gln Glu Gly Pro Ser Pro

```
145                 150                 155                 160

Glu Cys Leu Gly Gly Glu Glu Asp Glu Leu Pro Gly Leu Gly Gly Ala
                165                 170                 175

Pro Leu Gln Gly Leu Thr Leu Pro Asn Lys Ala Thr Leu Gly His Phe
            180                 185                 190

Tyr Glu Arg Cys Ile Arg Gly Ala Thr Ala Asp Ala Ala Arg Thr Arg
        195                 200                 205

Glu Phe Leu Glu Gly Phe Val Asp Asp Leu Leu Glu Ala Leu Arg Ser
    210                 215                 220

Leu Cys Asn Arg Asp Thr Asp Met Glu Val Glu Asp Phe Ile Gly Val
225                 230                 235                 240

Asp Ser Met Tyr Glu Asn Trp Gln Val Asp Arg Pro Leu Leu Cys His
                245                 250                 255

Leu Phe Val Pro Phe Thr Pro Pro Glu Pro Tyr Arg Phe His Pro Glu
            260                 265                 270

Leu Trp Cys Ser Gly Arg Ser Val Pro Leu Asp Arg Gln Gly Tyr Gly
        275                 280                 285

Gln Ile Lys Val Val Arg Ala Asp Gly Asp Thr Leu Ser Cys Ile Cys
    290                 295                 300

Gly Lys Thr Lys Leu Gly Glu Asp Met Leu Cys Leu Leu His Gly Arg
305                 310                 315                 320

Asn Ser Met Ala Pro Pro Cys Gly Asp Met Glu Asn Leu Leu Cys Ala
                325                 330                 335

Thr Asp Ser Leu Tyr Leu Asp Thr Met Gln Val Met Lys Trp Phe Gln
            340                 345                 350

Thr Ala Leu Thr Arg Ala Trp Lys Gly Ile Ala His Lys Tyr Glu Phe
        355                 360                 365

Asp Leu Ala Phe Gly Gln Leu Asp Ser Pro Gly Ser Leu Lys Ile Lys
    370                 375                 380
```

Phe Arg Ser Gly Lys Phe Met Pro Phe Asn Leu Ile Pro Val Ile Gln
385 390 395 400

Cys Asp Asp Ser Asp Leu Tyr Phe Val Ser His Leu Pro Arg Glu Pro
405 410 415

Ser Glu Gly Thr Pro Ala Ser Ser Thr Asp Trp Leu Leu Ser Phe Ala
420 425 430

Val Tyr Glu Arg His Phe Leu Arg Thr Thr Leu Lys Ala Leu Pro Glu
435 440 445

Gly Ala Cys His Leu Ser Cys Leu Gln Ile Ala Ser Phe Leu Leu Ser
450 455 460

Lys Gln Ser Arg Leu Thr Gly Pro Ser Gly Leu Ser Ser Tyr His Leu
465 470 475 480

Lys Thr Ala Leu Leu His Leu Leu Leu Leu Arg Gln Ala Ala Asp Trp
485 490 495

Lys Ala Gly Gln Leu Asp Ala Arg Leu His Glu Leu Leu Cys Phe Leu
500 505 510

Glu Lys Ser Leu Leu Gln Lys Lys Leu His His Phe Phe Ile Gly Asn
515 520 525

Arg Lys Val Pro Glu Ala Met Gly Leu Pro Glu Ala Val Leu Arg Ala
530 535 540

Glu Pro Leu Asn Leu Phe Arg Pro Phe Val Leu Gln Arg Ser Leu Tyr
545 550 555 560

Arg Lys Thr Leu Asp Ser Phe Tyr Glu Met Leu Lys Asn Ala Pro Ala
565 570 575

Leu Ile Ser Glu Tyr Ser Leu His Val Pro Ser Asp Gln Pro Thr Pro
580 585 590

Lys Ser

<210> 201
<211> 38
<212> PRT

<213> Homo sapien

<400> 201

Met Ser Leu His Ala Glu Val Gly Gly Ala Leu Lys Pro Val Ile Tyr
1 5 10 15

Ala Val Lys Thr Lys Trp Val Cys Tyr Leu Ile Ser Trp Gly Ile His
20 25 30

Gly Leu Ala Val Pro Gly
35

<210> 202
<211> 16
<212> PRT
<213> Homo sapien

<400> 202

Met Glu Arg Ile Gly Thr Phe Tyr Ser Gly Asn Thr Gln Pro Ala Thr
1 5 10 15

<210> 203
<211> 87
<212> PRT
<213> Homo sapien

<400> 203

Met Ala Glu Gly Val Gly Ala Gly Thr Leu Glu Ala Pro Pro Leu Leu
1 5 10 15

Ser Leu Pro Ser Ala Ser Pro Val Pro Pro Ala Ala Leu Val Thr Val
20 25 30

Ser Asp Gly Tyr Leu Pro Gly Phe Val Ala Ser Leu Ser Val Phe Ser
35 40 45

Cys Ser Asp Pro Leu Ala Gly Trp Leu Arg Lys Lys Lys Met Cys Phe
50 55 60

Arg Cys His Cys Asn Pro Gly His Gln Gly Asn Pro Ser Phe Pro Phe
65 70 75 80

Leu Ile Cys Ser Pro Arg Thr
85

<210> 204

<211> 252
<212> PRT
<213> Homo sapien

<400> 204

Met Ser Ile Tyr Lys Glu Pro Pro Pro Gly Met Phe Val Val Pro Asp
1 5 10 15

Thr Val Asp Met Thr Lys Ile His Ala Leu Ile Thr Gly Pro Phe Asp
20 25 30

Thr Pro Tyr Glu Gly Gly Phe Phe Leu Phe Val Phe Arg Cys Pro Pro
35 40 45

Asp Tyr Pro Ile His Pro Pro Arg Val Lys Leu Met Thr Thr Gly Asn
50 55 60

Asn Thr Val Arg Phe Asn Pro Asn Phe Tyr Arg Asn Gly Lys Val Cys
65 70 75 80

Leu Ser Ile Leu Gly Thr Trp Thr Gly Pro Ala Trp Ser Pro Ala Gln
85 90 95

Ser Ile Ser Ser Val Leu Ile Ser Ile Gln Ser Leu Met Thr Glu Asn
100 105 110

Pro Tyr His Asn Glu Pro Gly Phe Glu Gln Glu Arg His Pro Gly Asp
115 120 125

Ser Lys Asn Tyr Asn Glu Cys Ile Arg His Glu Thr Ile Arg Val Ala
130 135 140

Val Cys Asp Met Met Glu Gly Lys Cys Pro Cys Pro Glu Pro Leu Arg
145 150 155 160

Gly Val Met Glu Lys Ser Phe Leu Glu Tyr Tyr Asp Phe Tyr Glu Val
165 170 175

Ala Cys Lys Asp Arg Leu His Leu Gln Gly Gln Thr Met Gln Asp Pro
180 185 190

Phe Gly Glu Lys Arg Gly His Phe Asp Tyr Gln Ser Leu Leu Met Arg
195 200 205

Leu Gly Leu Ile Arg Gln Lys Val Leu Glu Arg Leu His Asn Glu Asn

210 215 220

Ala Glu Met Asp Ser Asp Ser Ser Ser Ser Gly Thr Glu Thr Asp Leu
225 230 235 240

His Gly Ser Leu Arg Val His Gly Ser Leu Arg Val
245 250

<210> 205
<211> 91
<212> PRT
<213> Homo sapien

<400> 205

Met Leu Thr Pro Ala Arg Pro Ser Cys His Thr Leu Ser Gly Arg Ser
1 5 10 15

Met Ala Tyr Arg Met Lys Arg Gly Thr Arg Asn Pro Cys Gly Arg Gly
20 25 30

Leu Asp Leu Lys Gln Cys Pro Leu Trp Leu Leu Leu Pro Trp Leu Thr
35 40 45

Gly Phe Leu Asp His Val His Phe Thr Gly Pro Trp Asp Leu His Leu
50 55 60

Leu Ala Ser Pro Ala Gly Leu Ile Pro Ala Arg Ala Pro Ser Phe Leu
65 70 75 80

Leu Met Val Phe Arg Trp Pro Asp His Gly Lys
85 90

<210> 206
<211> 213
<212> PRT
<213> Homo sapien

<400> 206

Ser Pro His Gln Ala Ala Ala Pro Val Asp Gln Thr Pro Arg Thr Leu
1 5 10 15

Ala Thr Met Gly Gln Arg Ala Leu Pro Ser Ser Leu Ala Leu Leu Ser
20 25 30

Arg Pro Leu Ser Pro Pro Pro Ala Ala Cys Ser Gly Asp Pro Gly Cys
35 40 45

Gly Ser Gly Ala Gly Leu Pro Ser Ala Ser Ala Ala Ala Gly Ile Ala
50 55 60

Ser Ser Ala Val Glu Ala Val Cys Gly Asp Ala Ala Pro Ala Cys Leu
65 70 75 80

Leu Arg Thr Pro Leu Arg Gly Leu Leu Lys Pro Thr Gly Pro Arg Ser
85 90 95

Thr Met Glu Cys Pro Pro Ala Leu Ile Val His Pro Pro Thr Gly Gly
100 105 110

Met Ala Arg Arg Ala Ala Ser Gln Pro Trp Ala Ala Ala Ser Ala Thr
115 120 125

Pro Met Leu Ser Ser Lys Ala Ser Leu Cys Ile Pro Thr Glu Arg Pro
130 135 140

Pro Pro Gln Pro Leu Met Arg Thr Pro Ala Ala Arg Ser His Trp Pro
145 150 155 160

Ile Pro His Pro Ala Ser Thr Ala Cys Pro Ala Pro Leu Pro Val Val
165 170 175

Leu Val Ala Pro Arg Ser Thr Ile Leu Ser Met Ser Arg Thr Trp Thr
180 185 190

Cys Arg Arg Trp Ala Val Ala Pro Cys Arg Ala Glu Lys Leu Met Cys
195 200 205

Ser Ser Ser Arg Ser
210

<210> 207
<211> 92
<212> PRT
<213> Homo sapien

<400> 207

Met Tyr Lys Gly Ala Ala Trp Arg Gly Lys Glu His Asn Lys Thr Pro
1 5 10 15

Leu Glu Val Phe Gln Arg Val Val Ser Gln Ile Ser Leu Ile Gln Glu
20 25 30

Glu Asp Asp Glu Arg Glu Arg Thr Trp Asn Tyr Leu Lys Ser Ser Asn
35 40 45

Ser Leu Val Leu Phe Asn Lys Lys Glu Phe Trp Phe Val Ala Glu Ser
50 55 60

Asp Leu Thr Ala Ala Asn Ser Ser Leu Leu Leu Arg Cys Ile Ser Asn
65 70 75 80

Ser Lys Leu Asp Ala Pro Pro Ser Leu Phe Phe Pro
85 90

<210> 208
<211> 130
<212> PRT
<213> Homo sapien

<400> 208

Met Val Cys Glu Asp Ala Pro Ser Phe Gln Met Ala Trp Glu Ser Gln
1 5 10 15

Met Ala Trp Glu Arg Gly Pro Ala Leu Leu Cys Cys Val Leu Ser Ala
20 25 30

Ser Gln Leu Ser Ser Gln Asp Gln Asp Pro Leu Gly His Ile Lys Ser
35 40 45

Leu Leu Tyr Pro Phe Gly Phe Pro Val Glu Leu Pro Arg Pro Gly Pro
50 55 60

Thr Gly Ala Tyr Lys Lys Val Lys Asn Gln Asn Gln Thr Thr Ser Ser
65 70 75 80

Glu Leu Leu Arg Lys Gln Thr Ser His Phe Asn Gln Arg Gly His Arg
85 90 95

Ala Arg Ser Lys Leu Leu Ala Ser Arg Gln Ile Pro Asp Arg Thr Phe
100 105 110

Lys Cys Gly Lys Trp Leu Pro Gln Val Pro Ser Pro Val Val Pro Ser
115 120 125

Pro Val
130

<210> 209
<211> 63
<212> PRT
<213> Homo sapien

<400> 209

Met Asn Asp Tyr Gly Leu Gly Leu Gly Phe Ile Thr Asn Pro Ile Ile
1 5 10 15

Asp His Leu Phe Pro Ala Leu Gly Ile Thr Ala Lys Pro Asn Gly Ser
20 25 30

Phe Ser Ile Thr Ala Ser Tyr Asn Phe His Ile Phe Leu Leu Phe Leu
35 40 45

Thr Gly Leu Gln Val Leu Ser Asn Val Leu Lys Leu Phe Asn Val
50 55 60

<210> 210
<211> 451
<212> PRT
<213> Homo sapien

<400> 210

Ala Thr Lys Thr Lys Ala Pro Asp Asp Leu Val Ala Pro Val Val Lys
1 5 10 15

Lys Pro His Ile Tyr Tyr Gly Ser Leu Glu Glu Lys Glu Arg Glu Arg
20 25 30

Leu Ala Lys Gly Glu Ser Gly Ile Leu Gly Lys Asp Gly Leu Lys Ala
35 40 45

Gly Ile Glu Ala Gly Asn Ile Asn Ile Thr Ser Gly Glu Val Phe Glu
50 55 60

Ile Glu Glu His Ile Ser Glu Arg Gln Ala Glu Val Leu Ala Glu Phe
65 70 75 80

Glu Arg Arg Lys Arg Ala Arg Gln Ile Asn Val Ser Thr Asp Asp Ser
85 90 95

Glu Val Lys Ala Cys Leu Arg Ala Leu Gly Glu Pro Ile Thr Leu Phe
100 105 110

```
Gly Glu Gly Pro Ala Glu Arg Arg Glu Arg Leu Arg Asn Ile Leu Ser
        115                 120                 125

Val Val Gly Thr Asp Ala Leu Lys Lys Thr Lys Lys Asp Asp Glu Lys
    130                 135                 140

Ser Lys Lys Ser Lys Glu Glu Tyr Gln Gln Thr Trp Tyr His Glu Gly
145                 150                 155                 160

Pro Asn Ser Leu Lys Val Ala Arg Leu Trp Ile Ala Asn Tyr Ser Leu
                165                 170                 175

Pro Arg Ala Met Lys Arg Leu Glu Glu Ala Arg Leu His Lys Glu Ile
            180                 185                 190

Pro Glu Thr Thr Arg Thr Ser Gln Met Gln Glu Leu His Lys Ser Leu
        195                 200                 205

Arg Ser Leu Asn Asn Phe Cys Ser Gln Ile Gly Asp Asp Arg Pro Ile
    210                 215                 220

Ser Tyr Cys His Phe Ser Pro Asn Ser Lys Met Leu Ala Thr Ala Cys
225                 230                 235                 240

Cys Asp Glu Pro Val Ala Asp Ile Glu Gly His Thr Val Arg Val Ala
                245                 250                 255

Arg Val Met Trp His Pro Ser Gly Arg Phe Leu Gly Thr Thr Cys Tyr
            260                 265                 270

Asp Arg Ser Trp Arg Leu Trp Asp Leu Glu Ala Gln Glu Glu Ile Leu
        275                 280                 285

His Gln Glu Gly His Ser Met Gly Val Tyr Asp Ile Ala Phe His Gln
    290                 295                 300

Asp Gly Ser Leu Ala Gly Thr Gly Gly Leu Asp Ala Phe Gly Arg Val
305                 310                 315                 320

Trp Asp Leu Arg Thr Gly Arg Cys Ile Met Phe Leu Glu Gly His Leu
                325                 330                 335

Lys Glu Ile Tyr Gly Ile Asn Phe Ser Pro Asn Gly Tyr His Ile Ala
            340                 345                 350
```

Thr Gly Ser Gly Asp Asn Thr Cys Lys Val Trp Asp Leu Arg Gln Arg
355 360 365

Arg Cys Val Tyr Thr Ile Pro Ala His Gln Asn Leu Val Thr Gly Val
370 375 380

Lys Phe Glu Pro Ile His Gly Asn Phe Leu Leu Thr Gly Ala Tyr Asp
385 390 395 400

Asn Thr Ala Lys Ile Trp Thr His Pro Gly Trp Ser Pro Leu Lys Thr
405 410 415

Leu Ala Gly His Glu Gly Lys Val Met Gly Leu Asp Ile Ser Ser Asp
420 425 430

Gly Gln Leu Ile Ala Thr Cys Ser Tyr Asp Arg Thr Phe Lys Leu Trp
435 440 445

Met Ala Glu
450

<210> 211
<211> 34
<212> PRT
<213> Homo sapien

<400> 211

Met Glu Ala Gln Gly Cys His Asp Gly Ser Val Val Ile Arg Glu Gly
1 5 10 15

Ala Pro Phe Ile Leu Leu Pro Thr Pro Leu Leu Cys Pro Phe Leu Pro
20 25 30

Leu Ile

<210> 212
<211> 610
<212> PRT
<213> Homo sapien

<400> 212

Gly Lys Ala Phe Ile Thr Cys Arg Thr Leu Leu Asn His Lys Ser Ile
1 5 10 15

His Phe Gly Asp Lys Pro Tyr Lys Cys Asp Glu Cys Glu Lys Ser Phe
20 25 30

Asn Tyr Ser Ser Leu Leu Ile Gln His Lys Val Ile His Thr Gly Glu
35 40 45

Lys Pro Tyr Glu Cys Asp Glu Cys Gly Lys Ala Phe Arg Asn Ser Ser
50 55 60

Gly Leu Ile Val His Lys Arg Ile His Thr Gly Glu Lys Pro Tyr Lys
65 70 75 80

Cys Asp Val Cys Gly Lys Ala Phe Ser Tyr Ser Ser Gly Leu Ala Val
85 90 95

His Lys Ser Ile His Pro Gly Lys Lys Ala His Glu Cys Lys Glu Cys
100 105 110

Gly Lys Ser Phe Ser Tyr Asn Ser Leu Leu Leu Gln His Arg Thr Ile
115 120 125

His Thr Gly Glu Arg Pro Tyr Val Cys Asp Val Cys Gly Lys Thr Phe
130 135 140

Arg Asn Asn Ala Gly Leu Lys Val His Arg Arg Leu His Thr Gly Glu
145 150 155 160

Lys Pro Tyr Lys Cys Asp Val Cys Gly Lys Ala Tyr Ile Ser Arg Ser
165 170 175

Ser Leu Lys Asn His Lys Gly Ile His Leu Gly Glu Lys Pro Tyr Lys
180 185 190

Cys Ser Tyr Cys Glu Lys Ser Phe Asn Tyr Ser Ser Ala Leu Glu Gln
195 200 205

His Lys Arg Ile His Thr Arg Glu Lys Pro Phe Gly Cys Asp Glu Cys
210 215 220

Gly Lys Ala Phe Arg Asn Asn Ser Gly Leu Lys Val His Lys Arg Ile
225 230 235 240

His Thr Gly Glu Arg Pro Tyr Lys Cys Glu Glu Cys Gly Lys Ala Tyr
245 250 255

Ile Ser Leu Ser Ser Leu Ile Asn His Lys Ser Val His Pro Gly Glu
260 265 270

Lys Pro Phe Lys Cys Asp Glu Cys Glu Lys Ala Phe Ile Thr Tyr Arg
275 280 285

Thr Leu Thr Asn His Lys Lys Val His Leu Gly Glu Lys Pro Tyr Lys
290 295 300

Cys Asp Val Cys Glu Lys Ser Phe Asn Tyr Thr Ser Leu Leu Ser Gln
305 310 315 320

His Arg Arg Val His Thr Arg Glu Lys Pro Tyr Glu Cys Asp Arg Cys
325 330 335

Glu Lys Val Phe Arg Asn Asn Ser Ser Leu Lys Val His Lys Arg Ile
340 345 350

His Thr Gly Glu Arg Pro Tyr Glu Cys Asp Val Cys Gly Lys Ala Tyr
355 360 365

Ile Ser His Ser Ser Leu Ile Asn His Lys Ser Thr His Pro Gly Lys
370 375 380

Thr Pro His Thr Cys Asp Glu Cys Gly Lys Ala Phe Phe Ser Ser Arg
385 390 395 400

Thr Leu Ile Ser His Lys Arg Val His Leu Gly Glu Lys Pro Phe Lys
405 410 415

Cys Val Glu Cys Gly Lys Ser Phe Ser Tyr Ser Ser Leu Leu Ser Gln
420 425 430

His Lys Arg Ile His Thr Gly Glu Lys Pro Tyr Val Cys Asp Arg Cys
435 440 445

Gly Lys Ala Phe Arg Asn Ser Ser Gly Leu Thr Val His Lys Arg Ile
450 455 460

His Thr Gly Glu Lys Pro Tyr Glu Cys Asp Glu Cys Gly Lys Ala Tyr
465 470 475 480

Ile Ser His Ser Ser Leu Ile Asn His Lys Ser Val His Gln Gly Lys

485 490 495

Gln Pro Tyr Asn Cys Glu Cys Gly Lys Ser Phe Asn Tyr Arg Ser Val
500 505 510

Leu Asp Gln His Lys Arg Ile His Thr Gly Lys Lys Pro Tyr Arg Cys
515 520 525

Asn Glu Cys Ala His Ile Pro Asn Ala Thr Ala Asp Leu Met Lys Val
530 535 540

Asp His Glu Glu Glu Pro Gln Leu Ser Glu Pro Tyr Leu Ser Lys Gln
545 550 555 560

Lys Lys Leu Met Ala Lys Ile Leu Glu His Asp Asp Val Ser Tyr Leu
565 570 575

Lys Lys Ile Leu Gly Glu Leu Ala Met Val Leu Asp Gln Ile Glu Ala
580 585 590

Glu Leu Glu Lys Arg Lys Leu Glu Asn Glu Ala Leu Ser Gln Trp Lys
595 600 605

Glu Phe
610

<210> 213
<211> 47
<212> PRT
<213> Homo sapien

<400> 213

Met Cys Ala Lys Trp Gly Glu Ile Gly Ala Gly Lys Pro Ile Pro His
1 5 10 15

Arg Gly Pro Ala Leu Ala Pro Gly Ser Pro His Ala Phe Phe Val Phe
20 25 30

Phe Phe Phe Phe Ala Ser Asp Gln Phe Thr Thr Val Ser Trp Thr
35 40 45

<210> 214
<211> 25
<212> PRT
<213> Homo sapien

<400> 214

Met Glu Thr Pro Ser Leu Glu Gly Thr Pro Arg Lys Pro Cys His Gly
1 5 10 15

Leu Leu Ser Leu Ser Ser Leu Leu Leu
20 25

<210> 215
<211> 29
<212> PRT
<213> Homo sapien

<400> 215

Met Ser Ser Tyr Gly Met Gln Gly Thr Val Gly Ser Arg Val Ser Ile
1 5 10 15

Leu Pro Thr Arg Ala Gln Gly Gln Ala Gly Glu Val Arg
20 25

<210> 216
<211> 64
<212> PRT
<213> Homo sapien

<400> 216

Met Val Thr Leu Asp Leu Leu Glu Arg Ala Gln Cys Asp Gly Ser Trp
1 5 10 15

Ser Arg Arg Gly Thr Pro Leu Leu Phe Tyr Phe Phe Cys Lys Val Leu
20 25 30

Thr Leu Glu Gly Tyr Ser Ile Gln Ser Leu Asn Met Phe Phe Lys Arg
35 40 45

Asn Lys Glu Gln Ala Thr Ala Leu Leu Glu Ile Thr Asn Arg Phe Leu
50 55 60

<210> 217
<211> 50
<212> PRT
<213> Homo sapien

<400> 217

Met Glu Pro His Ile Met Lys Phe Asn Ser His Val Lys Thr Phe Cys
1 5 10 15

Ile Val Gly Cys Gln Lys Tyr Phe Pro Asn Phe Arg Leu Thr Cys Arg
20 25 30

Val Gly Asp Gly Leu Pro Pro Tyr Asn Phe Lys Phe Val Ser Gln Ser
35 40 45

Leu Ala
50

<210> 218
<211> 785
<212> PRT
<213> Homo sapien

<400> 218

Lys Ala Lys Ile Ser Trp Glu Ala Pro Val Glu Lys Lys Thr Glu Cys
1 5 10 15

Ile Gln Lys Gly Lys Asn Asn Gln Val Gly Ala Trp Thr Leu Leu Leu
20 25 30

Val Leu Pro Ser Pro Gln Asp Val Ser Ser His Ser Gly Pro Arg Ala
35 40 45

Leu Thr Asn Arg Thr Pro Phe Cys Pro Gln Thr Glu Cys Phe Asn Phe
50 55 60

Ile Arg Phe Leu Gln Pro Tyr Asn Ala Ser His Leu Tyr Val Cys Gly
65 70 75 80

Thr Tyr Ala Phe Gln Pro Lys Cys Thr Tyr Val Asn Met Leu Thr Phe
85 90 95

Thr Leu Glu His Gly Glu Phe Glu Asp Gly Lys Gly Lys Cys Pro Tyr
100 105 110

Asp Pro Ala Lys Gly His Ala Gly Leu Leu Val Asp Gly Glu Leu Tyr
115 120 125

Ser Ala Thr Leu Asn Asn Phe Leu Gly Thr Glu Pro Ile Ile Leu Arg
130 135 140

Asn Met Gly Pro His His Ser Met Lys Thr Glu Tyr Leu Ala Phe Trp
145 150 155 160

```
Leu Asn Glu Pro His Phe Val Gly Ser Ala Tyr Val Pro Glu Ser Val
                165                 170                 175

Gly Ser Phe Thr Gly Asp Asp Asp Lys Val Tyr Phe Phe Phe Arg Glu
            180                 185                 190

Arg Ala Val Glu Ser Asp Cys Tyr Ala Glu Gln Val Val Ala Arg Val
        195                 200                 205

Ala Arg Val Cys Lys Gly Asp Met Gly Gly Ala Arg Thr Leu Gln Arg
    210                 215                 220

Lys Trp Thr Thr Phe Leu Lys Ala Arg Leu Ala Cys Ser Ala Pro Asn
225                 230                 235                 240

Trp Gln Leu Tyr Phe Asn Gln Leu Gln Ala Met His Thr Leu Gln Asp
                245                 250                 255

Thr Ser Trp His Asn Thr Thr Phe Phe Gly Val Phe Gln Ala Gln Trp
            260                 265                 270

Gly Asp Met Tyr Leu Ser Ala Ile Cys Glu Tyr Gln Leu Glu Glu Ile
        275                 280                 285

Gln Arg Val Phe Glu Gly Pro Tyr Lys Glu Tyr His Glu Glu Ala Gln
    290                 295                 300

Lys Trp Asp Arg Tyr Thr Asp Pro Val Pro Ser Pro Arg Pro Gly Ser
305                 310                 315                 320

Cys Ile Asn Asn Trp His Arg Arg His Gly Tyr Thr Ser Ser Leu Glu
                325                 330                 335

Leu Pro Asp Asn Ile Leu Asn Phe Val Lys Lys His Pro Leu Met Glu
            340                 345                 350

Glu Gln Val Gly Pro Arg Trp Ser Arg Pro Leu Leu Val Lys Lys Gly
        355                 360                 365

Thr Asn Phe Thr His Leu Val Ala Asp Arg Val Thr Gly Leu Asp Gly
    370                 375                 380

Ala Thr Tyr Thr Val Leu Phe Ile Gly Thr Gly Asp Gly Trp Leu Leu
385                 390                 395                 400
```

```
Lys Ala Val Ser Leu Gly Pro Trp Val His Leu Ile Glu Glu Leu Gln
                405                 410                 415

Leu Phe Asp Gln Glu Pro Met Arg Ser Leu Val Leu Ser Gln Ser Lys
            420                 425                 430

Val Lys Leu Leu Phe Ala Gly Ser Arg Ser Gln Leu Val Gln Leu Pro
        435                 440                 445

Val Ala Asp Cys Met Lys Tyr Arg Ser Cys Ala Asp Cys Val Leu Ala
    450                 455                 460

Arg Asp Pro Tyr Cys Ala Trp Ser Val Asn Thr Ser Arg Cys Val Ala
465                 470                 475                 480

Val Gly Gly His Ser Gly Ser Leu Leu Ile Gln His Val Met Thr Ser
                485                 490                 495

Asp Thr Ser Gly Ile Cys Asn Leu Arg Gly Ser Lys Lys Val Arg Pro
            500                 505                 510

Thr Pro Lys Asn Ile Thr Val Val Ala Gly Thr Asp Leu Val Leu Pro
        515                 520                 525

Cys His Leu Ser Ser Asn Leu Ala His Ala Arg Trp Thr Phe Gly Gly
    530                 535                 540

Arg Asp Leu Pro Ala Glu Gln Pro Gly Ser Phe Leu Tyr Asp Ala Arg
545                 550                 555                 560

Leu Gln Ala Leu Val Val Met Ala Ala Gln Pro Arg His Ala Gly Ala
                565                 570                 575

Tyr His Cys Phe Ser Glu Glu Gln Gly Ala Arg Leu Ala Ala Glu Gly
            580                 585                 590

Tyr Leu Val Ala Val Val Ala Gly Pro Ser Val Thr Leu Glu Ala Arg
        595                 600                 605

Ala Pro Leu Glu Asn Leu Gly Leu Val Trp Leu Ala Val Val Ala Leu
    610                 615                 620

Gly Ala Val Cys Leu Val Leu Leu Leu Leu Val Leu Ser Leu Arg Arg
625                 630                 635                 640
```

Arg Leu Arg Glu Glu Leu Glu Lys Gly Ala Lys Ala Thr Glu Arg Thr
645 650 655

Leu Val Tyr Pro Leu Glu Leu Pro Lys Glu Pro Thr Ser Pro Pro Phe
660 665 670

Arg Pro Cys Pro Glu Pro Asp Glu Lys Leu Trp Asp Pro Val Gly Tyr
675 680 685

Tyr Tyr Ser Asp Gly Ser Leu Lys Ile Val Pro Gly His Ala Arg Cys
690 695 700

Gln Pro Gly Gly Gly Pro Pro Ser Pro Pro Pro Gly Ile Pro Gly Gln
705 710 715 720

Pro Leu Pro Ser Pro Thr Arg Leu His Leu Gly Gly Gly Arg Asn Ser
725 730 735

Asn Ala Asn Gly Tyr Val Arg Leu Gln Leu Gly Gly Glu Asp Arg Gly
740 745 750

Gly Leu Gly His Pro Leu Pro Glu Leu Ala Asp Glu Leu Arg Arg Lys
755 760 765

Leu Gln Gln Arg Gln Pro Leu Pro Asp Ser Asn Pro Glu Glu Ser Ser
770 775 780

Val
785

<210> 219
<211> 66
<212> PRT
<213> Homo sapien

<400> 219

Met Lys Met Arg Ala Lys Ile Leu His Gln Asn Gly Asn Asp Pro Ile
1 5 10 15

Ser Pro Val Lys Ala Glu Trp Val Glu Trp Gly Leu Arg Val Trp Ile
20 25 30

Gln Cys Phe Glu Leu His Ser Ser Arg Glu Ala Val Gln Lys Gly Gly
35 40 45

Ile Leu Gly Asn Leu Arg Lys Ile Val Gly Glu Thr Ser Phe Leu Leu
50 55 60

Val Ser
65

<210> 220
<211> 128
<212> PRT
<213> Homo sapien

<400> 220

Glu Val Glu Gly Arg Ser Ala Cys Met Ala Met Gly Leu Phe Phe Ile
1 5 10 15

Pro Phe Leu Asn Cys Thr Gln Gln Gln Trp Phe Leu Leu Gly Leu Leu
20 25 30

Lys Thr Ala Gly Ile Trp Glu Lys Glu His His Arg Leu Ser Gln His
35 40 45

Gly Asn Ile Asn Leu Ile Pro Glu Lys Gly Arg Ser Pro Gln Arg Tyr
50 55 60

Val Arg Phe Asn Ser Phe Ser Ser Gly Pro Gly Ser Ser Phe Ser Cys
65 70 75 80

Ser Gly Leu Asn Arg Asp Ala Leu Ile Ser Leu Gly Ile Leu Leu Leu
85 90 95

Val Leu Ser Leu Thr Ser Gly Ala Lys Ile Arg Arg Pro Glu Phe Gln
100 105 110

Ile Tyr Ser Val Thr Gln Ser Leu Leu Gln Ser Leu Arg Asp Val Val
115 120 125

<210> 221
<211> 64
<212> PRT
<213> Homo sapien

<400> 221

Met Gly Ile Leu Glu Pro Gln Asp Val Arg Ala Gly Arg Asp Ala Ile
1 5 10 15

```
Pro Val Tyr Thr Arg Gly Asn Ser Ser Arg Leu Trp Glu Gly Arg Arg
            20                  25                  30

Val Leu Val Thr Glu Arg Glu Leu Lys Leu Arg Ile Pro Glu Ser Arg
        35                  40                  45

Ser Cys Leu Pro Ser Ala Ile Phe Leu Pro Ile Asn Leu Cys Tyr Val
    50                  55                  60
```

<210> 222
<211> 105
<212> PRT
<213> Homo sapien

<400> 222

```
Cys Lys Leu Phe Gly Arg Val Gly Asp Ala Val Ser Phe Cys His Pro
1               5                   10                  15

Gly Trp Ser Ala Val Ala Arg Ser Gln Leu Thr Ala Thr Ser Ala Leu
            20                  25                  30

Gln Gly Ser Gly Asn Ser Ala Ser Val Ser Ala Val Ala Gly Ile Thr
        35                  40                  45

Gly Met Arg His His Thr Arg Leu Ile Phe Val Phe Leu Val Glu Thr
    50                  55                  60

Arg Phe His His Val Gly Gln Asp Gly Leu Glu Pro Leu Thr Ser Gly
65                  70                  75                  80

Asp Leu Pro Ile Ser Ala Ser Gln Ser Ala Gly Ile Thr Ser Val Ser
                85                  90                  95

His Arg Ala Arg Pro Ala Asn Phe Phe
            100                 105
```

<210> 223
<211> 109
<212> PRT
<213> Homo sapien

<400> 223

```
Met Met Trp Leu Ser Val Gly Gly Gly Gly Arg Glu Trp Ser Glu Met
1               5                   10                  15
```

Leu Gly Val Val Trp Trp Trp Gly Gly Val Gly Val Trp Val Gly Val
20 25 30

Gly Val Cys Gly Cys Val Trp Trp Val Val Val Gly Val Trp Trp Trp
35 40 45

Arg Cys Val Gly Cys Gly Cys Val Val Trp Trp Gly Gly Val Val Gly
50 55 60

Val Gly Gly Cys Trp Gly Gly Cys Val Cys Val Val Gly Val Cys Val
65 70 75 80

Cys Val Gly Gly Gly Val Val Gly Arg Val Val Gly Gly Ala Gly Val
85 90 95

Cys Gly Gly Arg Cys Gly Cys Cys Val Val Trp Trp Cys
100 105

<210> 224
<211> 196
<212> PRT
<213> Homo sapien

<400> 224

Thr Arg Pro Gln Ser His Thr Thr Thr Glu His Pro Pro Pro Pro Pro
1 5 10 15

Thr Thr Ile His Ile Thr Gln Thr Leu His Lys Lys Thr Asn Thr Thr
20 25 30

Asn Thr Gln Gln Lys Lys His Thr Asn Thr Gln Ile Thr Ile Thr Gln
35 40 45

Gln His Thr Pro Gln His Thr Thr Thr Pro Pro Thr Pro His His Ser
50 55 60

Thr Pro Pro His Asn Thr Thr Pro Ala Pro Pro Pro His Thr Pro Ala
65 70 75 80

Pro Pro Thr Thr Arg Pro Thr Thr Pro Pro Pro Thr His Thr His Thr
85 90 95

Pro Thr Thr His Thr His Pro Pro Gln His Pro Pro Thr Pro Thr Thr
100 105 110

Thr Thr Pro Pro His His Ala Pro Thr Pro His Thr Pro Pro Pro Thr
115 120 125

Thr Pro Pro Arg Pro Pro Thr Thr His Thr His Thr Pro Pro His Pro
130 135 140

Pro Thr Pro Pro Pro Leu Pro Thr Thr Thr Pro His Pro Thr Ser His
145 150 155 160

Ser Thr Leu Ser Pro His His Pro His Ser Thr Thr Ser Ser Leu Pro
165 170 175

Ser Thr His Asn Asn Ile Thr Asn Thr Pro Pro Ala His Thr Leu Thr
180 185 190

Pro His Thr Ser
195

<210> 225
<211> 92
<212> PRT
<213> Homo sapien

<400> 225

Met Thr Ser Leu Pro Glu Gly Pro Arg Ala Ser Glu Asp Gly Ala Thr
1 5 10 15

Pro Glu Ala Gly Gly Phe Thr Asn Ser Ser His Leu Tyr Arg Arg Pro
20 25 30

Ala Arg Cys Gln Ala Cys Trp Gln Ala Gln Gly Lys Ala His Ser Thr
35 40 45

Ser Arg His Gly Pro Cys Ser His Gly Ala Tyr Ser Leu Ala Arg Gln
50 55 60

Thr Arg Asn Lys Lys Leu Gln Ser Ser Val Glu Val Cys Arg Val Val
65 70 75 80

Gly Tyr Ser Asp Leu Ala Leu Tyr Thr His Phe Ala
85 90

<210> 226
<211> 42
<212> PRT
<213> Homo sapien

<400> 226

Met Lys Ile Tyr Gly Ser Val Phe Gln Asn Asp Glu Glu Phe Gln Asp
1 5 10 15

Gly Gly Ser Gly Lys Ile Leu Leu Gln Glu Lys Ser Val Leu Gly Pro
20 25 30

Met Cys Lys His Leu Leu Arg Asn Leu Glu
35 40

<210> 227
<211> 57
<212> PRT
<213> Homo sapien

<400> 227

Met Leu Ser Gln Arg Tyr Arg Lys Val Leu Leu Gly Pro Ser Val Thr
1 5 10 15

Leu Ser Phe His Ile Pro Thr Leu His Arg Pro Ser Leu Gln Leu Pro
20 25 30

Ala Pro Ala Pro His Cys Arg Ser Pro Gly Phe Cys Leu Glu Leu Asn
35 40 45

Glu Glu Met Gly Pro Leu Ala Leu Ala
50 55

<210> 228
<211> 205
<212> PRT
<213> Homo sapien

<400> 228

Gln Gln Gly Lys Leu Val Ala Asp Ser Ala Lys His Leu Gly Leu Lys
1 5 10 15

His Val Val Tyr Ser Gly Leu Glu Asn Val Lys Arg Leu Thr Asp Gly
20 25 30

Lys Leu Glu Val Pro His Phe Asp Ser Lys Gly Glu Val Glu Glu Tyr
35 40 45

Phe Trp Ser Ile Gly Ile Pro Met Thr Ser Val Arg Val Ala Ala Tyr
50 55 60

```
Phe Glu Asn Phe Leu Ala Ala Trp Arg Pro Val Lys Ala Ser Asp Gly
65                  70                  75                  80

Asp Tyr Tyr Thr Leu Ala Val Pro Met Gly Asp Val Pro Met Asp Gly
                85                  90                  95

Ile Ser Val Ala Asp Ile Gly Ala Ala Val Ser Ser Ile Phe Asn Ser
            100                 105                 110

Pro Glu Glu Phe Leu Gly Lys Ala Val Gly Leu Ser Ala Glu Ala Leu
        115                 120                 125

Thr Ile Gln Gln Tyr Ala Asp Val Leu Ser Lys Ala Leu Gly Lys Glu
    130                 135                 140

Val Arg Asp Ala Lys Ile Thr Pro Glu Ala Phe Glu Lys Leu Gly Phe
145                 150                 155                 160

Pro Ala Ala Lys Glu Ile Ala Asn Met Cys Arg Phe Tyr Glu Met Lys
                165                 170                 175

Pro Asp Arg Asp Val Asn Leu Thr His Gln Leu Asn Pro Lys Val Lys
            180                 185                 190

Ser Phe Ser Gln Phe Ile Ser Glu Asn Gln Gly Ala Phe
        195                 200                 205
```

<210> 229
<211> 46
<212> PRT
<213> Homo sapien

<400> 229

```
Met Lys Lys Lys Val Leu Ser Ile Ile Cys Ile Ile Gly Ile His Met
1               5                   10                  15

Ser Leu His Lys Met Phe Asn Leu Lys Glu Ile Pro Leu Ile Leu Tyr
            20                  25                  30

Val Leu Leu Ser Val Val Cys Phe Ser Phe Leu Ile Leu Ser
        35                  40                  45
```

<210> 230
<211> 53

<212> PRT
<213> Homo sapien

<400> 230

Val Ala Gln Ala Gly Val Gln Trp Arg Asn Ala Asn Ser Leu Gln Pro
1 5 10 15

Ala Pro Ser Trp Leu Lys Gln Ala Leu His Leu Ser Pro Leu Ser Ser
20 25 30

Ala His Tyr Arg His Thr Pro Pro His Pro Ala Asn Phe Phe Glu Phe
35 40 45

Leu Glu Thr Gly Phe
50

<210> 231
<211> 30
<212> PRT
<213> Homo sapien

<400> 231

Met Gly Gln Val Gly Val Arg Gly Pro Gly Glu Val Arg Ala Leu Ser
1 5 10 15

Ser Lys Leu Ser Tyr Cys His Val Phe Val Pro Arg Arg Asp
20 25 30

<210> 232
<211> 39
<212> PRT
<213> Homo sapien

<400> 232

Met Val Phe Leu Gly Glu Leu Lys Thr Phe Ser Leu Val Ser Val Asn
1 5 10 15

Gln Arg Ala Phe Ser Leu Phe Leu Leu Leu Ile Pro Ser Ser Pro Val
20 25 30

Asn Tyr Phe Ser Phe His Trp
35

<210> 233
<211> 107
<212> PRT
<213> Homo sapien

<400> 233

Phe Phe Phe Phe Leu Leu Leu Phe Cys Asp Ser Leu Ala Leu Ser Pro
1 5 10 15

Arg Leu Gln Cys Ser Gly Thr Ile Ser Ala His Cys Asn Leu Cys Leu
20 25 30

Leu Gly Ser Ser Asn Ser Pro Val Ser Ala Ser Trp Val Ala Gly Thr
35 40 45

Thr Gly Ala Cys His His Ala Trp Leu Thr Phe Val Phe Leu Val Glu
50 55 60

Thr Gly Phe His His Val Gly Gln Ala Gly Leu Glu Phe Leu Thr Ser
65 70 75 80

Gly Asp Pro Pro Ala Leu Ala Ser Gln Ser Ala Glu Ile Thr Gly Val
85 90 95

Ser His Arg Ala Trp Pro Val Cys Phe Phe Asn
100 105

<210> 234
<211> 57
<212> PRT
<213> Homo sapien

<400> 234

Met Cys Ile Ile Leu Ser Ala His Ala Val Leu Gln Ala Ser Val Pro
1 5 10 15

Leu Ala Val His Val Ser Pro His Ala Arg Ala Gly Pro Ser Trp Ser
20 25 30

Ala Leu Val Ser Lys Trp Val Tyr Ala Glu Ala Asp Phe Gln Ser Val
35 40 45

Ser Cys Pro Pro Ile Gln His Ser Arg
50 55

<210> 235
<211> 50
<212> PRT
<213> Homo sapien

<400> 235

Met Lys Val Pro Ala Tyr Ile Asn His Leu Ala Arg Trp Trp Glu Ile
1 5 10 15

Leu Cys Ser Ser Asn Val Leu Leu Val Leu Gly Arg Asp Gly Ala His
20 25 30

Ser Gly Ala Lys Glu Asp Lys Lys Ser Met Gln Asn Leu Ser Leu Leu
35 40 45

Met Ala
50

<210> 236
<211> 44
<212> PRT
<213> Homo sapien

<400> 236

Met His Asn Trp Asp Cys Trp Asn Gly Pro Arg His Thr Thr Ala Gly
1 5 10 15

His Cys His Gln Glu Gly Ala Cys Val Leu Glu Gly Ser Gly Gln His
20 25 30

Arg Leu Ala Asn Leu Glu Gly Ser Gln Arg Asp Ser
35 40

<210> 237
<211> 146
<212> PRT
<213> Homo sapien

<400> 237

Met Gly Ala Arg Val Pro His Ala Ala Asp Gly Pro Ser Gln Val Glu
1 5 10 15

Leu Pro Gly Val Gln Ser Gly Ser Pro Leu Ala Asp Leu Met Leu Ser
20 25 30

Asp Arg Trp Asp Lys Phe Phe Cys His Ser Ala Gly Leu Cys Pro Glu
35 40 45

Ala Ser Leu Leu Ala Gly Cys Ala His Ala Trp Glu Lys Ala Trp Ala
50 55 60

Val Asn Tyr Gly His Thr Cys Ser Leu Cys Gly His Cys Ser Pro Ala
65 70 75 80

Pro Ile Pro Ile Pro Pro His Pro Thr His Pro Asn Thr His Thr Pro
85 90 95

Arg Pro Gln Thr Pro Thr Pro Thr Thr Pro His Pro Pro Thr Pro Thr
100 105 110

Pro Pro His Pro Pro Gln His Pro His Pro Arg Pro Pro Pro Thr Ser
115 120 125

Thr His Pro Pro Thr His Asn Thr Pro His Thr Thr His His Gln His
130 135 140

His His
145

<210> 238
<211> 47
<212> PRT
<213> Homo sapien

<400> 238

Met Tyr Arg Gln Tyr Gly Pro Trp Cys Thr Asn Ala Ala Ser Gly Arg
1 5 10 15

Arg Asp Val Met Asp Gly Arg Gly Arg Gly Thr Phe Asn Pro Ser Ser
20 25 30

Pro Phe Pro Pro Ser Gly Ala Ser Tyr Glu Ile Ser Val His Phe
35 40 45

<210> 239
<211> 91
<212> PRT
<213> Homo sapien

<400> 239

Met Val Lys Ile Ser Phe Gln Pro Ala Val Ala Gly Ile Lys Gly Asp
1 5 10 15

Lys Ala Asp Lys Ala Ser Ala Ser Ala Pro Ala Pro Ala Ser Ala Thr
20 25 30

Glu Ile Leu Leu Thr Pro Ala Arg Glu Glu Gln Pro Pro Gln His Arg
35 40 45

Ser Lys Arg Gly Gly Ser Val Gly Gly Val Cys Tyr Leu Ser Met Gly
50 55 60

Met Val Val Leu Leu Met Gly Leu Val Phe Ala Ser Val Tyr Ile Tyr
65 70 75 80

Arg Tyr Phe Phe Leu Ala Gln Leu Ala Arg Asp
85 90

<210> 240
<211> 188
<212> PRT
<213> Homo sapien

<400> 240

Met Arg Leu Val Gly Gly Val Gly Ser Phe Arg Leu Gly Gly Val Gly
1 5 10 15

Cys Gly Gly Gly Gly Gly Gly Ala Gly Ala Gly Ser Trp Val Trp Met
20 25 30

Gly Gly Trp Gly Gly Gly Ala Gly Ala Leu Trp Val Ala Val Val Gly
35 40 45

Gly Ala Arg Trp Trp Gly Gly Ala Gly Trp Gly Ser Cys Gly Arg Val
50 55 60

Leu Val Gly Gly Arg Ala Val Val Val Gly Arg Val Gly Val Val Gly
65 70 75 80

Trp Gly Trp Trp Arg Val Val Val Ala Gly Cys Val Cys Gly Gly Gly
85 90 95

Trp Arg Trp Trp Arg Ala Gly Val Gly Gly Gly Gly Gly Ala Val Ser
100 105 110

Gly Pro Ser Gly Ala Gly Pro Gly Arg Arg Cys Ser Met Val Glu Arg
115 120 125

Arg Arg Gly His Val Gly Ser Gly Gly Trp Ala Gly Arg Pro Gly Val
130 135 140

Val Gly Val Trp Ala Arg Cys Val Leu Val Ala Gly Ala Val Trp Arg
145 150 155 160

Arg Gly Gly Ala Val Trp Glu Trp Arg Gly Leu Gly Cys Gly Ala Trp
165 170 175

Cys Val Gly Arg Ser Trp Gly Glu Cys Gly Gly Arg
180 185

<210> 241
<211> 110
<212> PRT
<213> Homo sapien

<400> 241

Met Lys Leu Thr Leu Ser Glu Val Lys Met Glu Val Ile Gly Val Pro
1 5 10 15

Trp Arg Asn Gly Ser His Cys Phe Ile Ser Ile Thr Pro Gln Leu Lys
20 25 30

Phe Thr Pro Val Ser Gly His Lys Asn Met Arg Lys Glu Pro Cys Cys
35 40 45

Phe His Lys Gly Asn His Ser Ser Leu Ser Pro Leu Leu Ile Asn Leu
50 55 60

Lys Ser Trp Thr Pro Ser Phe Leu His Trp Pro Arg Pro Thr Leu Thr
65 70 75 80

His Leu Glu Pro Leu Phe Arg Ala Glu Trp His Glu Tyr Val Tyr Leu
85 90 95

Gly Arg Asp Gln Ser Ile Thr Gln Arg Arg Leu Glu Gln His
100 105 110

<210> 242
<211> 102
<212> PRT
<213> Homo sapien

<400> 242

Met Pro Ser Leu Pro Thr Arg Ser Leu Leu Ser Pro Cys Val Leu Glu
1 5 10 15

Leu Glu Glu Leu Thr Cys Ala Leu Cys Thr Trp Ala Phe Leu Leu Leu

20 25 30

Cys Leu Ala Leu Val Ala Asp Cys Pro Gly Leu Arg Gln Val Ile Pro
35 40 45

Gly Lys Gln Val Phe Val Leu Phe Ser Met Ser Gly Gly Arg Phe Ile
50 55 60

Leu Leu Ser Val Ser Ser His Phe Pro Ile Pro Phe Lys Lys Leu Trp
65 70 75 80

Pro Ala Gln Gly Arg Ala Leu Ser Cys Cys Ile Thr Ala Glu Pro Thr
85 90 95

Cys Pro His Ala Leu Leu
100

<210> 243
<211> 86
<212> PRT
<213> Homo sapien

<400> 243

Leu Ala Val Ser Leu Cys His Gln Ala Gly Val Gln Trp Cys Asn Pro
1 5 10 15

Gly Ser Leu Gln Pro Pro Pro Pro Gly Phe Lys Arg Phe Phe Cys Leu
20 25 30

Cys Leu Pro Ser Ser Trp Gly Tyr Arg His Thr Pro Pro Arg Pro Ala
35 40 45

Asn Phe Cys Val Phe Gly Arg Asp Gly Val Ser Pro Cys Trp Pro Gly
50 55 60

Trp Ser Leu Ser Leu Asp Val Ile Cys Asp Pro Pro Arg Gln Pro Pro
65 70 75 80

Lys Val Leu Gly Leu Gln
85

<210> 244
<211> 53
<212> PRT
<213> Homo sapien

<400> 244

Met Leu Leu Pro Phe Ala Val Arg Gly Leu Leu Thr Met Ala Arg Gly
1 5 10 15

Asp Val Ser Glu Ile Gln Val Val Val Ala Ser Trp Ser Thr Gln Leu
20 25 30

Ala His Met Gln Glu Glu Gly Leu Trp Pro Leu Ser Arg Ala Gly Gly
35 40 45

Leu Leu Pro Gln Ala
50

<210> 245
<211> 183
<212> PRT
<213> Homo sapien

<400> 245

Leu Thr Pro Ala Gly Val Pro Trp Cys His Leu Gly Ser Leu Gln Pro
1 5 10 15

Leu Pro Pro Arg Phe Lys Ala Val Phe Ser Arg Leu Ala Pro Ser Leu
20 25 30

Glu Tyr Ala Trp Asp Tyr Arg Ala Pro Thr Ser His Ala Arg Leu Ile
35 40 45

Ser Leu Ala Phe Leu Val Glu Thr Gly Phe Ser Pro Thr Val Ala Arg
50 55 60

Leu Val Ser Asn Ser Trp Pro Pro Val Val Arg Pro Pro Leu Pro Ser
65 70 75 80

Gln Ser Ala Gly Ile Thr Gly Val Gly Pro Pro Cys Leu Ala Arg Pro
85 90 95

Ile Leu Pro Pro His Pro Phe Phe Phe Phe Phe Asp Met Glu Ser His
100 105 110

Ala Ile Thr Gln Ala Gly Val Gln Trp Arg His Leu Gly Ser Leu Gln
115 120 125

Pro Pro Pro Pro Met Phe Lys Ala Ser Ser Cys Leu Ser Leu Leu Ser
130 135 140

Ser Trp Asp Tyr Arg Arg Pro Pro Pro Arg Pro Ala Ile Phe Cys Ile
145 150 155 160

Phe Ser Arg Asp Gly Val Ser Pro Cys Ala Pro Gly Trp Ser Arg Ser
165 170 175

Pro Asp Leu Thr Pro Asp Leu
180

<210> 246
<211> 12
<212> PRT
<213> Homo sapien

<400> 246

Met Ala Pro Asp Thr Asn Thr Phe Leu His Pro Phe
1 5 10

<210> 247
<211> 240
<212> PRT
<213> Homo sapien

<400> 247

Met Gly Asn Cys Gln Ala Gly His Asn Leu His Leu Cys Leu Ala His
1 5 10 15

His Pro Pro Leu Val Cys Ala Thr Leu Ile Leu Leu Leu Leu Gly Leu
20 25 30

Ser Gly Leu Gly Leu Gly Ser Phe Leu Leu Thr His Arg Thr Gly Leu
35 40 45

Arg Ser Pro Asp Ile Pro Gln Asp Trp Val Ser Phe Leu Arg Ser Phe
50 55 60

Gly Gln Leu Thr Leu Cys Pro Arg Asn Gly Thr Val Thr Gly Lys Trp
65 70 75 80

Arg Gly Ser His Val Val Gly Leu Leu Thr Thr Leu Asn Phe Gly Asp
85 90 95

Gly Pro Asp Arg Asn Lys Thr Arg Thr Phe Gln Ala Thr Val Leu Gly
100 105 110

Ser Gln Met Gly Leu Lys Gly Ser Ser Ala Gly Gln Leu Val Leu Ile
115 120 125

Thr Ala Arg Val Thr Thr Glu Arg Thr Ala Gly Thr Cys Leu Tyr Phe
130 135 140

Ser Ala Val Pro Gly Ile Leu Pro Ser Ser Gln Pro Pro Ile Ser Cys
145 150 155 160

Ser Glu Glu Gly Ala Gly Asn Ala Thr Leu Ser Pro Arg Met Gly Glu
165 170 175

Glu Cys Val Ser Val Trp Ser His Glu Gly Leu Val Leu Thr Lys Leu
180 185 190

Leu Thr Ser Glu Glu Leu Ala Leu Cys Gly Ser Arg Leu Leu Val Leu
195 200 205

Gly Ser Phe Leu Leu Leu Phe Cys Gly Leu Leu Cys Cys Val Thr Ala
210 215 220

Met Cys Phe His Pro Arg Arg Glu Ser His Trp Ser Arg Thr Arg Leu
225 230 235 240

<210> 248
<211> 75
<212> PRT
<213> Homo sapien

<400> 248

Met Arg Arg Ala Val Ala Ser Val Met Tyr Arg Trp Ser Arg Pro Arg
1 5 10 15

Tyr Thr Gln Glu Ala Arg Arg Tyr Phe Phe Phe Ser Glu Leu Ser Pro
20 25 30

Gly Ser Lys Gly Glu Ala Met Gly Asp Pro Gly Met Val Leu Ala Ser
35 40 45

Gly Gly Cys Phe Leu Val Thr Gly Val Ser Ser Lys Gln Asn Gly Ile
50 55 60

Arg Met Lys Arg Gly Lys Gly Met Gly His Lys
65 70 75

<210> 249
<211> 594
<212> PRT
<213> Homo sapien

<400> 249

Val Pro Gly Arg Lys Leu His Arg Ser Arg Leu Gln Ala Ala Ala Pro
1 5 10 15

Arg Pro Ser Thr Cys Ala Gln Ser Leu Cys Trp Ser Arg Pro Pro Ala
20 25 30

Ala Gly Thr Gly Thr Gly Asp Pro Ser Gln Ser Lys Ala Pro Thr Met
35 40 45

Ala Met Gly Leu Phe Arg Val Cys Leu Val Val Val Thr Ala Ile Ile
50 55 60

Asn His Pro Leu Leu Phe Pro Arg Glu Asn Ala Thr Val Pro Glu Asn
65 70 75 80

Glu Glu Glu Ile Ile Arg Lys Met Gln Ala His Gln Glu Lys Leu Gln
85 90 95

Leu Glu Gln Leu Arg Leu Glu Glu Glu Val Ala Arg Leu Ala Ala Glu
100 105 110

Lys Glu Ala Leu Glu Gln Val Ala Glu Glu Gly Arg Gln Gln Asn Glu
115 120 125

Thr Arg Val Ala Trp Asp Leu Trp Ser Thr Leu Cys Met Ile Leu Phe
130 135 140

Leu Met Ile Glu Val Trp Arg Gln Asp His Gln Glu Gly Pro Ser Pro
145 150 155 160

Glu Cys Leu Gly Gly Glu Glu Asp Glu Leu Pro Gly Leu Gly Gly Ala
165 170 175

Pro Leu Gln Gly Leu Thr Leu Pro Asn Lys Ala Thr Leu Gly His Phe
180 185 190

Tyr Glu Arg Cys Ile Arg Gly Ala Thr Ala Asp Ala Ala Arg Thr Arg
195 200 205

```
Glu Phe Leu Glu Gly Phe Val Asp Asp Leu Leu Glu Ala Leu Arg Ser
    210                 215                 220

Leu Cys Asn Arg Asp Thr Asp Met Glu Val Glu Asp Phe Ile Gly Val
225                 230                 235                 240

Asp Ser Met Tyr Glu Asn Trp Gln Val Asp Arg Pro Leu Leu Cys His
                245                 250                 255

Leu Phe Val Pro Phe Thr Pro Pro Glu Pro Tyr Arg Phe His Pro Glu
            260                 265                 270

Leu Trp Cys Ser Gly Arg Ser Val Pro Leu Asp Arg Gln Gly Tyr Gly
        275                 280                 285

Gln Ile Lys Val Val Arg Ala Asp Gly Asp Thr Leu Ser Cys Ile Cys
    290                 295                 300

Gly Lys Thr Lys Leu Gly Glu Asp Met Leu Cys Leu Leu His Gly Arg
305                 310                 315                 320

Asn Ser Met Ala Pro Pro Cys Gly Asp Met Glu Asn Leu Leu Cys Ala
                325                 330                 335

Thr Asp Ser Leu Tyr Leu Asp Thr Met Gln Val Met Lys Trp Phe Gln
            340                 345                 350

Thr Ala Leu Thr Arg Ala Trp Lys Gly Ile Ala His Lys Tyr Glu Phe
        355                 360                 365

Asp Leu Ala Phe Gly Gln Leu Asp Ser Pro Gly Ser Leu Lys Ile Lys
    370                 375                 380

Phe Arg Ser Gly Lys Phe Met Pro Phe Asn Leu Ile Pro Val Ile Gln
385                 390                 395                 400

Cys Asp Asp Ser Asp Leu Tyr Phe Val Ser His Leu Pro Arg Glu Pro
                405                 410                 415

Ser Glu Gly Thr Pro Ala Ser Ser Thr Asp Trp Leu Leu Ser Phe Ala
            420                 425                 430

Val Tyr Glu Arg His Phe Leu Arg Thr Thr Leu Lys Ala Leu Pro Glu
        435                 440                 445
```

```
Gly Ala Cys His Leu Ser Cys Leu Gln Ile Ala Ser Phe Leu Leu Ser
    450                 455                 460

Lys Gln Ser Arg Leu Thr Gly Pro Ser Gly Leu Ser Ser Tyr His Leu
465                 470                 475                 480

Lys Thr Ala Leu Leu His Leu Leu Leu Leu Arg Gln Ala Ala Asp Trp
                485                 490                 495

Lys Ala Gly Gln Leu Asp Ala Arg Leu His Glu Leu Leu Cys Phe Leu
            500                 505                 510

Glu Lys Ser Leu Leu Gln Lys Lys Leu His His Phe Phe Ile Gly Asn
        515                 520                 525

Arg Lys Val Pro Glu Ala Met Gly Leu Pro Glu Ala Val Leu Arg Ala
    530                 535                 540

Glu Pro Leu Asn Leu Phe Arg Pro Phe Val Leu Gln Arg Ser Leu Tyr
545                 550                 555                 560

Arg Lys Thr Leu Asp Ser Phe Tyr Glu Met Leu Lys Asn Ala Pro Ala
                565                 570                 575

Leu Ile Ser Glu Tyr Ser Leu His Val Pro Ser Asp Gln Pro Thr Pro
            580                 585                 590

Lys Ser
```

<210> 250
<211> 23
<212> PRT
<213> Homo sapien

<400> 250

```
Met Tyr Cys Ile Gly Gly Trp Ala Gly Pro Thr Leu Cys Tyr Val Lys
1               5                   10                  15

Glu Leu Val Leu Val Leu Gly
            20
```

<210> 251
<211> 213

<212> PRT
<213> Homo sapien

<400> 251

Ser Pro His Gln Ala Ala Ala Pro Val Asp Gln Thr Pro Arg Thr Leu
1 5 10 15

Ala Thr Met Gly Gln Arg Ala Leu Pro Ser Ser Leu Ala Leu Leu Ser
20 25 30

Arg Pro Leu Ser Pro Pro Pro Ala Ala Cys Ser Gly Asp Pro Gly Cys
35 40 45

Gly Ser Gly Ala Gly Leu Pro Ser Ala Ser Ala Ala Ala Gly Ile Ala
50 55 60

Ser Ser Ala Val Glu Ala Val Cys Gly Asp Ala Ala Pro Ala Cys Leu
65 70 75 80

Leu Arg Thr Pro Leu Arg Gly Leu Leu Lys Pro Thr Gly Pro Arg Ser
85 90 95

Thr Met Glu Cys Pro Pro Ala Leu Ile Val His Pro Pro Thr Gly Gly
100 105 110

Met Ala Arg Arg Ala Ala Ser Gln Pro Trp Ala Ala Ala Ser Ala Thr
115 120 125

Pro Met Leu Ser Ser Lys Ala Ser Leu Cys Ile Pro Thr Glu Arg Pro
130 135 140

Pro Pro Gln Pro Leu Met Arg Thr Pro Ala Ala Arg Ser His Trp Pro
145 150 155 160

Ile Pro His Pro Ala Ser Thr Ala Cys Pro Ala Pro Leu Pro Val Val
165 170 175

Leu Val Ala Pro Arg Ser Thr Ile Leu Ser Met Ser Arg Thr Trp Thr
180 185 190

Cys Arg Arg Trp Ala Val Ala Pro Cys Arg Ala Glu Lys Leu Met Cys
195 200 205

Ser Ser Ser Arg Ser
210

<210> 252
<211> 32
<212> PRT
<213> Homo sapien

<400> 252

Met His Glu Leu Thr Ala Arg Leu Thr Gln Pro Leu Asn Ser Gly Ser
1 5 10 15

Cys Phe Ser Leu Ala Ala Ile His His Met Arg Arg Arg Ser Met His
20 25 30

<210> 253
<211> 58
<212> PRT
<213> Homo sapien

<400> 253

Met Ser Leu Gln Leu Gln Ile Leu Asn Val Ser Pro Val Ile Trp His
1 5 10 15

Phe Arg His Ser Tyr Leu Lys Pro Gln Phe Ser Leu Pro Val Lys Trp
20 25 30

Gly Ile Ile Ile Pro Ile Leu Pro Arg Leu Leu Lys Gly Leu Ser Glu
35 40 45

Leu Ile Cys Lys Met Leu Asn Arg Thr Gln
50 55

<210> 254
<211> 34
<212> PRT
<213> Homo sapien

<400> 254

Met Gly Ser Ala Phe Val Leu Leu Ser Trp Arg Ala Cys Leu Cys Cys
1 5 10 15

Arg Ala Val Ser Val Val Gly Ile Ala Leu Leu Phe Pro Ala Thr Gly
20 25 30

Gln Ile

<210> 255
<211> 74
<212> PRT
<213> Homo sapien

<400> 255

Lys Arg Phe Phe Phe Phe Pro Ala Pro Ile Phe Cys Lys Thr Glu Val
1 5 10 15

Pro Glu His Arg Arg Ser Ser Ser Gln Ala Asn Phe Ile Lys Lys Lys
20 25 30

Leu Glu Val Cys Phe Asp Phe Ala Val Ile Cys Phe Ile Thr Ser Ile
35 40 45

Phe Gly Glu Gln Pro Gln Leu Leu Ile Phe Met Glu Lys Tyr Phe Gln
50 55 60

Val Gln Gly Gln Tyr Ile Ser Gln Ser Glu
65 70

<210> 256
<211> 34
<212> PRT
<213> Homo sapien

<400> 256

Met Ile Lys Val Cys Val Pro Ile Thr Phe Pro Leu Pro Glu Arg Arg
1 5 10 15

Val Ser Arg Lys Ile Asn Ser Ile Leu Asp Ala Gly Thr Ser Pro Arg
20 25 30

Pro Arg

<210> 257
<211> 37
<212> PRT
<213> Homo sapien

<400> 257

Met Asn Ser Ser Asn Arg Arg Leu Phe Trp Lys Lys Ser Gln Gly Leu
1 5 10 15

Ser Pro Ser Trp Val Ala Pro Tyr Lys Ser Asn Ser Ser Ser Gly Ser
20 25 30

Leu Val Tyr Pro Leu
35

<210> 258
<211> 73
<212> PRT
<213> Homo sapien

<400> 258

Met Glu Phe Leu Leu Leu Glu Val Glu Lys Tyr Asn Ile Ile Lys Lys
1 5 10 15

Asp Val Ile Pro Thr Arg Gly Leu Arg Gly Lys Leu Lys Asp Ile Lys
20 25 30

Gln Ser Asn Leu Val Ile Val Lys Thr Ile Tyr Val Gly His Arg Thr
35 40 45

Glu Asp Gln Val Ser Lys Glu Asp Gly Ser Val Pro Phe Val Ser Pro
50 55 60

Val Pro Lys Ala Val Phe Gly Ala Ser
65 70

<210> 259
<211> 1533
<212> PRT
<213> Homo sapien

<400> 259

Met Tyr Ile Arg Val Ser Tyr Asp Thr Lys Pro Asp Ser Leu Leu His
1 5 10 15

Leu Met Val Lys Asp Trp Gln Leu Glu Leu Pro Lys Leu Leu Ile Ser
20 25 30

Val His Gly Gly Leu Gln Asn Phe Glu Met Gln Pro Lys Leu Lys Gln
35 40 45

Val Phe Gly Lys Gly Leu Ile Lys Ala Ala Met Thr Thr Gly Ala Trp
50 55 60

Ile Phe Thr Gly Gly Val Ser Thr Gly Val Ile Ser His Val Gly Asp
65 70 75 80

```
Ala Leu Lys Asp His Ser Ser Lys Ser Arg Gly Arg Val Cys Ala Ile
                85                  90                  95

Gly Ile Ala Pro Trp Gly Ile Val Glu Asn Lys Glu Asp Leu Val Gly
            100                 105                 110

Lys Asp Val Thr Arg Val Tyr Gln Thr Met Ser Asn Pro Leu Ser Lys
        115                 120                 125

Leu Ser Val Leu Asn Asn Ser His Thr His Phe Ile Leu Ala Asp Asn
    130                 135                 140

Gly Thr Leu Gly Lys Tyr Gly Ala Glu Val Lys Leu Arg Arg Leu Leu
145                 150                 155                 160

Glu Lys His Ile Ser Leu Gln Lys Ile Asn Thr Arg Leu Gly Gln Gly
                165                 170                 175

Val Pro Leu Val Gly Leu Val Val Glu Gly Gly Pro Asn Val Val Ser
            180                 185                 190

Ile Val Leu Glu Tyr Leu Gln Glu Glu Pro Pro Ile Pro Val Val Ile
        195                 200                 205

Cys Asp Gly Ser Gly Arg Ala Ser Asp Ile Leu Ser Phe Ala His Lys
    210                 215                 220

Tyr Cys Glu Glu Gly Gly Ile Ile Asn Glu Ser Leu Arg Glu Gln Leu
225                 230                 235                 240

Leu Val Thr Ile Gln Lys Thr Phe Asn Tyr Asn Lys Ala Gln Ser His
                245                 250                 255

Gln Leu Phe Ala Ile Ile Met Glu Cys Met Lys Lys Lys Glu Leu Val
            260                 265                 270

Thr Val Phe Arg Met Gly Ser Glu Gly Gln Gln Asp Ile Glu Met Ala
        275                 280                 285

Ile Leu Thr Ala Leu Leu Lys Gly Thr Asn Val Ser Ala Pro Asp Gln
    290                 295                 300

Leu Ser Leu Ala Leu Ala Trp Asn Arg Val Asp Ile Ala Arg Ser Gln
305                 310                 315                 320
```

```
Ile Phe Val Phe Gly Pro His Trp Thr Pro Leu Gly Ser Leu Ala Pro
                325                 330                 335

Pro Thr Asp Ser Lys Ala Thr Glu Lys Glu Lys Lys Pro Pro Met Ala
            340                 345                 350

Thr Thr Lys Gly Gly Arg Gly Lys Gly Lys Gly Lys Lys Lys Gly Lys
        355                 360                 365

Val Lys Glu Glu Val Glu Glu Glu Thr Asp Pro Arg Lys Ile Glu Leu
    370                 375                 380

Leu Asn Trp Val Asn Ala Leu Glu Gln Ala Met Leu Asp Ala Leu Val
385                 390                 395                 400

Leu Asp Arg Val Asp Phe Val Lys Leu Leu Ile Glu Asn Gly Val Asn
                405                 410                 415

Met Gln His Phe Leu Thr Ile Pro Arg Leu Glu Glu Leu Tyr Asn Thr
            420                 425                 430

Arg Leu Gly Pro Pro Asn Thr Leu His Leu Leu Val Arg Asp Val Lys
        435                 440                 445

Lys Ser Asn Leu Pro Pro Asp Tyr His Ile Ser Leu Ile Asp Ile Gly
    450                 455                 460

Leu Val Leu Glu Tyr Leu Met Gly Gly Ala Tyr Arg Cys Asn Tyr Thr
465                 470                 475                 480

Arg Lys Asn Phe Arg Thr Leu Tyr Asn Asn Leu Phe Gly Pro Lys Arg
                485                 490                 495

Pro Lys Ala Leu Lys Leu Leu Gly Met Glu Asp Asp Glu Pro Pro Ala
            500                 505                 510

Lys Gly Lys Lys Lys Lys Lys Lys Lys Lys Glu Glu Glu Ile Asp Ile
        515                 520                 525

Asp Val Asp Asp Pro Ala Val Ser Arg Phe Gln Tyr Pro Phe His Glu
    530                 535                 540

Leu Met Val Trp Ala Val Leu Met Lys Arg Gln Lys Met Ala Val Phe
```

545 550 555 560

Leu Trp Gln Arg Gly Glu Glu Ser Met Ala Lys Ala Leu Val Ala Cys
565 570 575

Lys Leu Tyr Lys Ala Met Ala His Glu Ser Ser Glu Ser Asp Leu Val
580 585 590

Asp Asp Ile Ser Gln Asp Leu Asp Asn Asn Ser Lys Asp Phe Gly Gln
595 600 605

Leu Ala Leu Glu Leu Leu Asp Gln Ser Tyr Lys His Asp Glu Gln Ile
610 615 620

Ala Met Lys Leu Leu Thr Tyr Glu Leu Lys Asn Trp Ser Asn Ser Thr
625 630 635 640

Cys Leu Lys Leu Ala Val Ala Ala Lys His Arg Asp Phe Ile Ala His
645 650 655

Thr Cys Ser Gln Met Leu Leu Thr Asp Met Trp Met Gly Arg Leu Arg
660 665 670

Met Arg Lys Asn Pro Gly Leu Lys Val Ile Met Gly Ile Leu Leu Pro
675 680 685

Pro Thr Ile Leu Phe Leu Glu Phe Arg Thr Tyr Asp Asp Phe Ser Tyr
690 695 700

Gln Thr Ser Lys Glu Asn Glu Asp Gly Lys Glu Lys Glu Glu Glu Asn
705 710 715 720

Thr Asp Ala Asn Ala Asp Ala Gly Ser Arg Lys Gly Asp Glu Glu Asn
725 730 735

Glu His Lys Lys Gln Arg Ser Ile Pro Ile Gly Thr Lys Ile Cys Glu
740 745 750

Phe Tyr Asn Ala Pro Ile Val Lys Phe Trp Phe Tyr Thr Ile Ser Tyr
755 760 765

Leu Gly Tyr Leu Leu Leu Phe Asn Tyr Val Ile Leu Val Arg Met Asp
770 775 780

```
Gly Trp Pro Ser Leu Gln Glu Trp Ile Val Ile Ser Tyr Ile Val Ser
785                 790                 795                 800

Leu Ala Leu Glu Lys Ile Arg Glu Ile Leu Met Ser Glu Pro Gly Lys
                805                 810                 815

Leu Ser Gln Lys Ile Lys Val Trp Leu Gln Glu Tyr Trp Asn Ile Thr
            820                 825                 830

Asp Leu Val Ala Ile Ser Thr Phe Met Ile Gly Ala Ile Leu Arg Leu
        835                 840                 845

Gln Asn Gln Pro Tyr Met Gly Tyr Gly Arg Val Ile Tyr Cys Val Asp
    850                 855                 860

Ile Ile Phe Trp Tyr Ile Arg Val Leu Asp Ile Phe Gly Val Asn Lys
865                 870                 875                 880

Tyr Leu Gly Pro Tyr Val Met Met Ile Gly Lys Met Met Ile Asp Met
                885                 890                 895

Leu Tyr Phe Val Val Ile Met Leu Val Val Leu Met Ser Phe Gly Val
            900                 905                 910

Ala Arg Gln Ala Ile Leu His Pro Glu Glu Lys Pro Ser Trp Lys Leu
        915                 920                 925

Ala Arg Asn Ile Phe Tyr Met Pro Tyr Trp Met Ile Tyr Gly Glu Val
    930                 935                 940

Phe Ala Asp Gln Ile Asp Leu Tyr Ala Met Glu Ile Asn Pro Pro Cys
945                 950                 955                 960

Gly Glu Asn Leu Tyr Asp Glu Glu Gly Lys Arg Leu Pro Pro Cys Ile
                965                 970                 975

Pro Gly Ala Trp Leu Thr Pro Ala Leu Met Ala Cys Tyr Leu Leu Val
            980                 985                 990

Ala Asn Ile Leu Leu Val Asn Leu  Leu Ile Ala Val Phe  Asn Asn Thr
        995                 1000                1005

Phe Phe  Glu Val Lys Ser Ile  Ser Asn Gln Val Trp  Lys Phe Gln
    1010                1015                1020
```

```
Arg Tyr  Gln Leu Ile Met Thr  Phe His Asp Arg Pro  Val Leu Pro
    1025                 1030                 1035

Pro Pro  Met Ile Ile Leu Ser  His Ile Tyr Ile Ile  Ile Met Arg
    1040                 1045                 1050

Leu Ser  Gly Arg Cys Arg Lys  Lys Arg Glu Gly Asp  Gln Glu Glu
    1055                 1060                 1065

Arg Asp  Arg Gly Leu Lys Leu  Phe Leu Ser Asp Glu  Glu Leu Lys
    1070                 1075                 1080

Arg Leu  His Glu Phe Glu Glu  Gln Cys Val Gln Glu  His Phe Arg
    1085                 1090                 1095

Glu Lys  Glu Asp Glu Gln Gln  Ser Ser Ser Asp Glu  Arg Ile Arg
    1100                 1105                 1110

Val Thr  Ser Glu Arg Val Glu  Asn Met Ser Met Arg  Leu Glu Glu
    1115                 1120                 1125

Ile Asn  Glu Arg Glu Thr Phe  Met Lys Thr Ser Leu  Gln Thr Val
    1130                 1135                 1140

Asp Leu  Arg Leu Ala Gln Leu  Glu Glu Leu Ser Asn  Arg Met Val
    1145                 1150                 1155

Asn Ala  Leu Glu Asn Leu Ala  Gly Ile Asp Arg Ser  Asp Leu Ile
    1160                 1165                 1170

Gln Ala  Arg Ser Arg Ala Ser  Ser Glu Cys Glu Ala  Thr Tyr Leu
    1175                 1180                 1185

Leu Arg  Gln Ser Ser Ile Asn  Ser Ala Asp Gly Tyr  Ser Leu Tyr
    1190                 1195                 1200

Arg Tyr  His Phe Asn Gly Glu  Glu Leu Leu Phe Glu  Asp Thr Ser
    1205                 1210                 1215

Leu Ser  Thr Ser Pro Gly Thr  Gly Val Arg Lys Lys  Thr Cys Ser
    1220                 1225                 1230

Phe Arg  Ile Lys Glu Glu Lys  Asp Val Lys Thr His  Leu Val Pro
    1235                 1240                 1245
```

```
Glu Cys  Gln Asn Ser Leu His  Leu Ser Leu Gly Thr  Ser Thr Ser
    1250                 1255                 1260

Ala Thr  Pro Asp Gly Ser His  Leu Ala Val Asp Asp  Leu Lys Asn
    1265                 1270                 1275

Ala Glu  Glu Ser Lys Leu Gly  Pro Asp Ile Gly Ile  Ser Lys Glu
    1280                 1285                 1290

Asp Asp  Glu Arg Gln Thr Asp  Ser Lys Lys Glu Glu  Thr Ile Ser
    1295                 1300                 1305

Pro Ser  Leu Asn Lys Thr Asp  Val Ile His Gly Gln  Asp Lys Ser
    1310                 1315                 1320

Asp Val  Gln Asn Thr Gln Leu  Thr Val Glu Thr Thr  Asn Ile Glu
    1325                 1330                 1335

Gly Thr  Ile Ser Tyr Pro Leu  Glu Glu Thr Lys Ile  Thr Arg Tyr
    1340                 1345                 1350

Phe Pro  Asp Glu Thr Ile Asn  Ala Cys Lys Thr Met  Lys Ser Arg
    1355                 1360                 1365

Ser Phe  Val Tyr Ser Arg Gly  Arg Lys Leu Val Gly  Gly Val Asn
    1370                 1375                 1380

Gln Asp  Val Glu Tyr Ser Ser  Ile Thr Asp Gln Gln  Leu Thr Thr
    1385                 1390                 1395

Glu Trp  Gln Cys Gln Val Gln  Lys Ile Thr Arg Ser  His Ser Thr
    1400                 1405                 1410

Asp Ile  Pro Tyr Ile Val Ser  Glu Ala Ala Val Gln  Ala Glu Gln
    1415                 1420                 1425

Lys Glu  Gln Phe Ala Asp Met  Gln Asp Glu His His  Val Ala Glu
    1430                 1435                 1440

Ala Ile  Pro Arg Ile Pro Arg  Leu Ser Leu Thr Ile  Thr Asp Arg
    1445                 1450                 1455

Asn Gly  Met Glu Asn Leu Leu  Ser Val Lys Pro Asp  Gln Thr Leu
```

1460 1465 1470

Gly Phe Pro Ser Leu Arg Ser Lys Ser Leu His Gly His Pro Arg
1475 1480 1485

Asn Val Lys Ser Ile Gln Gly Lys Leu Asp Arg Ser Gly His Ala
1490 1495 1500

Ser Ser Val Ser Ser Leu Val Ile Val Ser Gly Met Thr Ala Glu
1505 1510 1515

Glu Lys Lys Val Lys Lys Glu Lys Ala Ser Thr Glu Thr Glu Cys
1520 1525 1530

<210> 260
<211> 92
<212> PRT
<213> Homo sapien

<400> 260

Met Ile Ile Leu Val Val Gly Arg Ile Thr Arg Gly Asn Ala Leu Tyr
1 5 10 15

Ser Gln Glu Glu Cys Cys Val Cys Thr Thr Gln Leu Thr Thr Trp Val
20 25 30

Val Cys Ser Thr Leu His Cys Val Ser Ile Leu Trp Ser Val Arg Pro
35 40 45

Ser Leu Ser Glu Gly Gly Tyr Leu Pro Leu Ala Ala Ser Val Ser Ala
50 55 60

Ala Ile Val Val Cys Phe Val Cys Val Cys Val Val Ser Cys His Asp
65 70 75 80

Ala Thr Ile Leu Leu Arg Ile Gly Asn Phe Gly Gly
85 90

<210> 261
<211> 66
<212> PRT
<213> Homo sapien

<400> 261

Met Glu Leu Leu Thr Asp Lys Gly Glu Ile Leu Asp Leu Glu Pro Phe
1 5 10 15

Pro Ala Ile Leu Leu Phe Ser Leu Cys Leu Gly Ser Trp Phe His Ser
20 25 30

Ala Arg His Glu Gly Pro Phe Gln Phe Asp Asp Ile Arg Leu Leu Thr
35 40 45

Leu Ser Trp Met Pro Cys Cys Leu Gln Gln His Asp Phe Thr Val Cys
50 55 60

Phe Ser
65

<210> 262
<211> 90
<212> PRT
<213> Homo sapien

<400> 262

Met Trp Asn Ile Pro Gly Leu Ala Gly Ala Met Pro Ala Met Gln Thr
1 5 10 15

Ser Pro Glu Pro Ser His Pro Gly Ser Val Arg Val Pro Arg Ala Val
20 25 30

Ala Pro His Pro Pro Pro Thr Gly Pro Cys Ser Trp Ser Cys Cys Asp
35 40 45

Ser Phe Ile Ile Pro Trp Ala Gly Val Gly Leu Ser Leu Cys Phe Cys
50 55 60

Leu Leu Phe Lys Glu Asp Glu Val Ser Met Glu Asn Lys Thr Asn Val
65 70 75 80

Val Thr Pro Ser Leu Arg Arg Val His Cys
85 90

<210> 263
<211> 13
<212> PRT
<213> Homo sapien

<400> 263

Met Ser Gly Gln Pro Arg Pro Thr Ser Pro Cys Val Leu
1 5 10

<210> 264
<211> 100
<212> PRT
<213> Homo sapien

<400> 264

Phe Phe Leu Arg Trp Ser Leu Ala Gln Val Ala Gln Ala Ala Arg Gln
1 5 10 15

Trp Leu Asn Leu Ser Ser Leu Gln Pro Pro Pro Pro Gly Phe Lys Arg
20 25 30

Phe Ser Cys Leu Gly Leu Leu Ser Ser Trp Asp Tyr Arg His Ala Pro
35 40 45

Pro Arg Pro Ala Ile Phe Val Phe Leu Val Glu Met Gly Phe His His
50 55 60

Ile Val Gln Ala Gly Leu Lys Pro Leu Thr Ser Gly Asp Leu Ala Thr
65 70 75 80

Ser Ala Phe Gln Ser Ala Glu Ile Ile Gly Val Ser His Cys Ala Gln
85 90 95

Pro Gln Lys Ser
100

<210> 265
<211> 10
<212> PRT
<213> Homo sapien

<400> 265

Met Lys Gly Lys Ile Leu Ile Phe Pro Ile
1 5 10

<210> 266
<211> 43
<212> PRT
<213> Homo sapien

<400> 266

Met Ser Pro Glu Pro Ser His Phe Ser Pro Pro Ala Pro Pro Ser Phe
1 5 10 15

Ser Pro Thr His Pro Ser Leu Pro Leu Thr Trp Ile Ser Ala Pro Ala

20 25 30

Ala Ser Pro Leu Pro Leu Leu Leu Pro Thr Phe
35 40

<210> 267
<211> 124
<212> PRT
<213> Homo sapien

<400> 267

Met Val Phe Tyr Cys Ile Leu Phe Leu Gln Leu Ile Gln Phe Cys Met
1 5 10 15

Ser Phe Leu Ser Phe Leu Gly Glu Asn Ile Leu Cys Gln Leu Phe Ser
20 25 30

Thr Val Leu His Tyr Ile Leu Lys Gln Gly Cys Gln Leu Glu Thr Gln
35 40 45

Pro Ser Asp Tyr Lys Ala Gln Asn Val Thr Phe Asn Cys Ala Pro Pro
50 55 60

Gly Gly Leu Ala Leu Gly Lys Asp Gly Glu Arg Asn Ile Leu Arg Tyr
65 70 75 80

Glu His Phe Leu Phe Cys Leu Gln Cys Cys Asp Leu Val Gln Gln Leu
85 90 95

Gln Asn Cys Ser His Leu Asn Arg Cys Ser Phe Ser Phe Phe Thr Leu
100 105 110

Leu Tyr Lys Arg Leu Val Ser Gln Leu His Tyr His
115 120

<210> 268
<211> 67
<212> PRT
<213> Homo sapien

<400> 268

Met Pro Glu Phe His Pro His Ser Leu Glu Leu Phe Thr Tyr Ser Pro
1 5 10 15

Ser Gln Glu Leu Leu Asp Glu His Gln Glu Met Arg Phe Lys Tyr Asn
20 25 30

Thr Glu Lys Cys Ala Gln Ala Gly Tyr His Pro Cys Trp Asn Leu Ala
35 40 45

Leu Ala Asn Trp Ala Thr Arg Val Pro Ala Arg Ala Asp Pro Ser Gln
50 55 60

Ser Ala Gly
65

<210> 269
<211> 23
<212> PRT
<213> Homo sapien

<400> 269

Met Thr Asp Leu Lys Glu Asn Ser Lys Ala Asp Leu Glu Asn Leu Leu
1 5 10 15

Leu Phe Leu Ser Pro Asn Pro
20

<210> 270
<211> 46
<212> PRT
<213> Homo sapien

<400> 270

Met Glu Asn Leu Ser Ser Ile Ser Glu Val Val Asn Ala Ile Ser Gly
1 5 10 15

Ile Gln Arg Leu Ala Val Lys Ser Ser Leu Gly Ser Leu Tyr Leu Thr
20 25 30

Phe Phe Leu Val Ser Ile Leu Lys Met Gln Ser His Ile Leu
35 40 45

<210> 271
<211> 15
<212> PRT
<213> Homo sapien

<400> 271

Met Thr Glu Glu Gly Glu Ser Leu Ser Gly Gln Ser Leu Gly Trp
1 5 10 15

<210> 272
<211> 46
<212> PRT
<213> Homo sapien

<400> 272

Met Pro Ser Ala Arg Met Ser Asp Gly Leu Val Ala Ala Glu Val Gln
1 5 10 15

Ser Pro Val Ile Phe Leu Phe Gly Pro Ile Trp Leu Leu Ile Leu Met
20 25 30

His Gln Asn Phe Met Tyr Asn His Met Asp Leu Tyr Val Asn
35 40 45

<210> 273
<211> 32
<212> PRT
<213> Homo sapien

<400> 273

Met Gly Arg Ala Leu Pro Leu Ser Ala Ala Pro Ser Leu Ser Leu Cys
1 5 10 15

Leu Pro Ala Gln Lys Arg Trp Leu Trp Pro Arg Gly Ser Gly Arg Asp
20 25 30

<210> 274
<211> 224
<212> PRT
<213> Homo sapien

<400> 274

Met Ala Val Gly Asn Ile Asn Glu Leu Pro Glu Asn Ile Leu Leu Glu
1 5 10 15

Leu Phe Thr His Val Pro Ala Arg Gln Leu Leu Leu Asn Cys Arg Leu
20 25 30

Val Cys Ser Leu Trp Arg Asp Leu Ile Asp Leu Val Thr Leu Trp Lys
35 40 45

Arg Lys Cys Leu Arg Glu Gly Phe Ile Thr Glu Asp Trp Asp Gln Pro
50 55 60

Val Ala Asp Trp Lys Ile Phe Tyr Phe Leu Arg Ser Leu His Arg Asn
65 70 75 80

Leu Leu His Asn Pro Cys Ala Glu Glu Gly Phe Glu Phe Trp Ser Leu
85 90 95

Asp Val Asn Gly Gly Asp Glu Trp Lys Val Glu Asp Leu Ser Arg Asp
100 105 110

Gln Arg Lys Glu Phe Pro Asn Asp Gln Val Arg Ser Gln Ala Arg Leu
115 120 125

Arg Val Gln Val Pro Ala Val Arg Ser Ala Pro Val Val Arg Ala Arg
130 135 140

Ala Ser Gly Asp Leu Pro Ala Arg Pro Gly Asp His Pro Ala Glu Glu
145 150 155 160

Arg Cys Gln Val Glu Gly Gly Leu Pro His Ile Leu Gln Leu Pro Ala
165 170 175

Arg Arg Pro Leu His Leu Val Ser Ala Arg Arg Arg Gly His Ser Leu
180 185 190

Leu Gly Arg Leu Val Arg Pro Glu Gly His Gln Gln Gln His His His
195 200 205

Arg Ala Pro Ala Ala Leu Thr Pro Pro Glu Pro Pro Ser Ala Glu Pro
210 215 220

<210> 275
<211> 33
<212> PRT
<213> Homo sapien

<400> 275

Met Gly Gly Gln Ala Thr Arg Tyr Tyr Ile Ile Asn Ile Leu Ser Gly
1 5 10 15

Lys Ile Ser Leu Phe Arg Ala Ile Arg Gln Val Ala Lys Asn Phe Ile
20 25 30

Leu

<210> 276
<211> 77

<212> PRT
<213> Homo sapien

<400> 276

Met Asn Gly Lys Thr Lys Val Glu Arg Asn Ile Leu Ser Tyr Ile Ile
1 5 10 15

Leu Gln Ile Lys Thr Phe Lys Asn Gln Ile Val Phe Leu Val Leu Arg
20 25 30

Thr Asn Arg Lys Cys Leu Ile Ile Tyr Phe Ile Ser Thr Arg Gln Lys
35 40 45

Tyr Ser Tyr Ala Ala Asp Val Arg Glu Gly Gly Glu Phe Pro Gln Pro
50 55 60

Ser Met Lys Lys Asp Lys Gly Pro Tyr Pro Leu Ala Val
65 70 75

<210> 277
<211> 39
<212> PRT
<213> Homo sapien

<400> 277

Met Tyr Val Arg Ser Ile His Leu Lys Ser Met Val Gln Ile Ala Lys
1 5 10 15

Ile Gly Pro Gly Glu Thr Cys Ser His Phe Leu Lys Thr Cys Thr Ser
20 25 30

Ala Ala Asn His Ala Thr Pro
35

<210> 278
<211> 26
<212> PRT
<213> Homo sapien

<400> 278

Met Pro Ile Arg Leu Cys Val Cys Ala Arg Phe Leu Lys Thr Ala Asn
1 5 10 15

Tyr Ile Val Ser Ser Gln Met Ser Gly Phe
20 25

<210> 279
<211> 149
<212> PRT
<213> Homo sapien

<400> 279

Met Leu Val Phe Ser Ala Gly Arg Leu Ala Cys Trp Arg Ala Val Cys
1 5 10 15

Trp Leu Gly Arg Cys Ala Cys Ala Ser Ser Arg Val Cys Leu Arg Leu
20 25 30

Val Leu Ser Trp Ser Arg Val Val Cys Phe Trp Trp Ser Phe Trp Leu
35 40 45

Phe Val Ser Val Val Cys Phe Val Phe Ser Cys Phe Val Ser Leu Leu
50 55 60

Cys Cys Cys Gly Val Arg Leu Tyr Phe Val Val Ser Trp Gly Val Phe
65 70 75 80

Phe Cys Asp Leu Leu Arg Cys Cys Tyr Asp Asn Val Cys Phe Ala His
85 90 95

Pro Thr Val Cys Phe Ser Ser Cys Pro Phe Phe Gly Val Leu Asn Tyr
100 105 110

Val Phe Phe Ile Leu Phe Pro His Trp Gly Val Cys Val Gly Gly Val
115 120 125

Val Pro Phe Ala Ala Val Phe Ser Gly Phe Phe Trp Ser Cys Pro Cys
130 135 140

Phe Val Ala Ala Arg
145

<210> 280
<211> 54
<212> PRT
<213> Homo sapien

<400> 280

Met Ile Leu Lys Gly Thr Leu Thr Ile Tyr Asn Lys Ser Phe Gln Tyr
1 5 10 15

Tyr Ser Ser Ser Leu Thr Ser Glu Ser Leu Val Tyr Val Ile Leu Ser

20 25 30

Arg Lys Lys Thr Thr Tyr Lys Ser His Phe Pro Thr Lys Leu Ile Gln
35 40 45

His Pro Thr Leu Lys Ile
50

<210> 281
<211> 114
<212> PRT
<213> Homo sapien

<400> 281

Val Ala Gly Ile Thr Gly Ile His His His Thr Gln Leu Phe Phe Cys
1 5 10 15

Ile Phe Val Arg Asp Arg Phe Leu His Val Gly Gln Ala Gly Leu Glu
20 25 30

Leu Pro Thr Ser Gly Asp Pro Pro Thr Ser Ala Ser Gln Ser Asp Asp
35 40 45

Phe Ile Phe Ile Phe Asn Cys Ile Asn Leu His Leu Asp Asn Asp Phe
50 55 60

Val Lys Gly Val Cys Cys Val Gln Asn Leu Arg Tyr Trp Leu Arg Val
65 70 75 80

Lys Tyr Ile Ile Phe Ile Ile Cys Trp Val Ala Ser Ser Tyr Ala Ala
85 90 95

Phe Phe Leu Ser Thr Phe Ile Lys Ser Ser Phe Leu Lys Leu Phe Ile
100 105 110

Ile Phe

<210> 282
<211> 171
<212> PRT
<213> Homo sapien

<400> 282

Met Leu Phe Cys Ile Phe Thr Val Tyr Cys Phe Tyr Asn Lys Tyr Lys
1 5 10 15

Met Lys Met Phe Met Leu Thr Lys Arg Thr Lys Asn Asn Lys Gln Gln
20 25 30

Lys Thr Lys Gly Trp Gly Cys His Thr Cys Gly Pro Lys Ala Gly Phe
35 40 45

Pro Gly Gly Gly His Leu Val Leu Ser Arg Pro His Asn Ser Pro Pro
50 55 60

Lys Tyr Tyr Arg Glu Thr Thr Gly Arg Thr Thr Gln His Thr Lys Arg
65 70 75 80

His Asn Thr Gln Asn His His Thr Ala Thr Pro Ala His Arg Arg Gln
85 90 95

Arg Thr Arg Arg Glu Gln Lys Glu Lys Gly Gln Gln Lys Lys Ala Ser
100 105 110

Ser Thr Ile Thr Thr Gln Ser His Asp Lys Lys Arg Arg Thr Met Thr
115 120 125

Lys Thr Ser Ser Ser Thr Arg His Arg Gln Asp Lys Ser Lys Lys Asp
130 135 140

Arg Thr Arg Gln Lys Thr Thr Arg Asp Glu Thr Thr Lys Lys Pro His
145 150 155 160

Lys Lys Ala Ser Glu Asn Lys Asn Gln Leu Thr
165 170

<210> 283
<211> 90
<212> PRT
<213> Homo sapien

<400> 283

Met Asn Ala Thr Val Leu Ser Val Phe Lys Ala Lys Leu Leu Trp Lys
1 5 10 15

Leu Gly Gly Gly Pro Pro Cys Gly Pro Pro Ala Ala Leu Cys Leu Pro
20 25 30

Leu Gly Ala Pro Glu Leu Met Pro Val Val Ile Ser Ala Met Leu Asp
35 40 45

```
Ala Arg Ser Gln Arg Ser Ala Ser Leu Ser Gln Leu Ala Cys Ala Ala
    50                  55                  60

Leu Thr Trp Leu Pro Ala Val Leu Arg Asn Leu His Trp Trp Asp Lys
65                  70                  75                  80

Gly Met Lys Arg Ile Asn Lys Asp Leu Lys
                85                  90
```

<210> 284
<211> 154
<212> PRT
<213> Homo sapien

<400> 284

```
Lys Glu Ala Pro Ser Ser Gln Asp Ile Leu Val Phe Leu Thr Gly Gln
1               5                   10                  15

Glu Glu Ile Glu Ala Met Ser Lys Thr Cys Arg Asp Ile Ala Lys His
            20                  25                  30

Leu Pro Asp Gly Cys Pro Ala Met Leu Val Leu Pro Leu Tyr Ala Ser
        35                  40                  45

Leu Pro Tyr Ala Gln Gln Leu Arg Val Phe Gln Gly Ala Pro Lys Gly
    50                  55                  60

Tyr Arg Lys Val Ile Ile Ser Thr Asn Ile Ala Glu Thr Ser Ile Thr
65                  70                  75                  80

Ile Thr Gly Ile Lys Tyr Val Val Asp Thr Gly Met Val Lys Ala Lys
                85                  90                  95

Lys Tyr Asn Pro Asp Ser Gly Leu Glu Val Leu Ala Val Gln Arg Val
            100                 105                 110

Ser Lys Thr Gln Ala Trp Gln Arg Thr Gly Arg Ala Gly Arg Glu Asp
        115                 120                 125

Ser Gly Ile Cys Tyr Arg Leu Tyr Thr Glu Asp Glu Phe Glu Lys Phe
    130                 135                 140

Asp Lys Met Thr Val Pro Glu Ile Gln Arg
145                 150
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068633 A3

(51) International Patent Classification[7]: C12N 15/09, C12Q 1/68, G01N 33/574, C07K 14/47

(21) International Application Number: PCT/US01/43612

(22) International Filing Date:
21 November 2001 (21.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/252,500 22 November 2000 (22.11.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/252,500 (CIP)
Filed on 22 November 2000 (22.11.2000)

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **CHEN, Sei-Yu** [—/US]; 160 Mira Street, Foster City, CA 94404 (US). **SUN, Yongming** [CN/US]; 551 Shoal Drive, Redwood City, CA 94065 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way, #14, San Jose, CA 95117 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
27 March 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/068633 A3

(54) Title: COMPOSITIONS AND METHODS RELATING TO LUNG SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic lung cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating lung cancer and non-cancerous disease states in lung, identifying lung tissue, monitoring and identifying and/or designing and antagonists of polypeptide of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered lung tissue for treatment and research.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/43612

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/09 C12Q1/68 G01N33/574 C07K14/47

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q G01N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, EMBL, BIOSIS, MEDLINE, PAJ, WPI Data, SEQUENCE SEARCH

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 2 July 2000 (2000-07-02) DIAS NETO E. ET AL.: "QV0-HT0368-090200-099-d10 HT0368 Homo sapiens cDNA, mRNA sequence." retrieved from EBI Database accession no. BE156557 XP002204032 99.3% identity (100% ungapped) in 309 nt overlap (338-31:20-327) with SEQ ID NO:1. abstract -& EMANUEL DIAZ NETO ET AL: "SHOTGUN SEQUENCING OF THE HUMAN TRANSCRIPTOME WITH ORF EXPRESSED SEQUENCE TAGS" PNAS, vol. 97, no. 7, 28 March 2000 (2000-03-28), pages 3491-3496, XP000996193 the whole document -/--	1,2,4-8, 15

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 August 2002	07 11 2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Schmitz, T

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 16 February 2000 (2000-02-16) PHILLIMORE, B.: "Human DNA sequence from clone RP11-427L15 on chromosome 10" retrieved from EBI Database accession no. AL139407 XP002204033 99.3% identity (100% ungapped) in 309 nt overlap (338-31:11807-12114) with SEQ ID NO:1. abstract	1,2,4-8
X,P	WO 01 22920 A (HUMAN GENOME SCIENCES INC ;ROSEN CRAIG A (US); BARASH STEVEN C (US) 5 April 2001 (2001-04-05) SEQ ID NO:969: 99.3% identity (100% ungapped) in 309 nt overlap (31-338:827-1134) with SEQ ID NO:1.	1-17
A	WO 00 08206 A (MACINA ROBERTO A ;SUN YONGMING (CN); YANG FEI (CN); RECIPON HERVE) 17 February 2000 (2000-02-17) the whole document	
A	WO 99 60160 A (DIADEXUS LLC ;SUN YONGMING (US); YANG FEI (US); MACINA ROBERTO A () 25 November 1999 (1999-11-25) the whole document	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/43612

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claim 16 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II *Observations where unity of invention is lacking* (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-17 all partially

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: claims 1-17 (all partially)

A nucleic acid molecule as defined by SEQ ID NO:1.

Further vectors and host cells comprising said nucleic acid sequence. Polypeptides encoded by said nucleic acid sequence, a method of producing said polypeptide, an antibody binding to said polypeptide. Methods for determining the presence of said nucleic acid or polypeptide in a sample. A kit, a vaccine and methods of treatment and diagnosis involving said nucleic acid or polypeptide.

Invention 2: claims 1-17 (all partially)

As subject 1, but referring to SEQ ID NO:2

Invention 3: claims 1-17 (all partially)

As subject 1, but referring to SEQ ID NO:3

.
.
.

Invention 164: claims 1-17 (all partially)

As subject 1, but referring to SEQ ID NO:164

INTERNATIONAL SEARCH REPORT

Information on patent family members

National Application No
PCT/US 01/43612

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0122920 A	05-04-2001	AU 7721500 A	30-04-2001
		WO 0122920 A2	05-04-2001
WO 0008206 A	17-02-2000	EP 1104486 A1	06-06-2001
		JP 2002522046 T	23-07-2002
		WO 0008206 A1	17-02-2000
WO 9960160 A	25-11-1999	CA 2328138 A1	25-11-1999
		EP 1082459 A1	14-03-2001
		JP 2002515262 T	28-05-2002
		WO 9960160 A1	25-11-1999